

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 11:56:36 ; Search time 10354 Seconds
(without alignments)
11513.488 Million cell updates/sec

Title: US-09-843-007A-1
Perfect score: 2914
Sequence: 1 GAGTTTTCGCTCCGAACC.....CGTTTCAGACGGCATTTCGC 2914

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank

- 1: gb_ba.*
- 2: gb_hvg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2641	90.6	2883	6	A47369	A47369 Sequence 1
2	2641	90.6	2883	6	A47785	A47785 Sequence 1
3	2093	71.8	2115	1	NFO011781	AJ011781 Neisseria
4	1921.4	65.9	1939	6	AX107013	AX107013 Sequence
5	1921.4	65.9	6878	6	AX107015	AX107015 Sequence
6	1891.4	64.9	1910	6	AX107014	AX107014 Sequence
7	1891.4	64.9	6851	6	AX107016	AX107016 Sequence
8	1853.4	63.6	1967	1	AY093334	AY093334 Neisseria
9	1847	63.4	1967	1	AY093335	AY093335 Neisseria
10	655	22.5	795	6	AX300255	AX300255 Sequence
11	655	22.5	329861	1	NMA522491	AL162756 Neisseria
12	639	21.9	765	6	AX565901	AX565901 Sequence
13	639	21.9	13001	1	AE002494	AE002494 Neisseria
14	639	21.9	349980	6	AX044032	AX044032 Sequence
15	380.2	13.0	10719	1	AE001946	AE001946 Deinococc
16	314	10.8	11605	1	AE012454	AE012454 Xanthomon
17	303.8	10.4	10261	1	AE011997	AE011997 Xanthomon
18	239	8.2	10822	1	AE005791	AE005791 Caulobact
19	170.6	5.9	10644	1	AE005392	AE005392 Escherich
20	170.6	5.9	10838	1	AE015192	AE015192 Shigella
21	170.6	5.9	257071	1	AP002558	AP002558 Escherich
22	170.6	5.9	290582	1	AE016984	AE016984 Shigella
23	169	5.8	759	1	ECARODG	X59503 E.coli aroD
24	169	5.8	759	1	SDU82289	U82289 Shigella dy
25	169	5.8	10685	1	AB000264	AB000264 Escherich
26	169	5.8	19521	1	D90811	D90811 E.coli geno
27	164.2	5.6	300475	1	AE016761	AE016761 Escherich
28	163.8	5.6	1798	1	ECAROD	X04306 E. coli aro
29	163	5.6	941	1	STAROD3D	X54546 S. typhi ar
30	161.4	5.5	233050	1	AL627271	AL627271 Salmonell
31	161.4	5.5	300523	1	AE016838	AE016838 Salmonell
32	157.8	5.4	759	1	AF184963	AF184963 Salmonell
33	157.8	5.4	20279	1	AE008759	AE008759 Salmonell
34	156.4	5.4	420	1	NGU39803	U39803 Neisseria g
35	153.2	5.3	249050	1	AL596165	AL596165 Listeria
36	153.2	5.3	349980	6	AX417038	AX417038 Sequence
37	153.2	5.3	349980	6	AX417041	AX417041 Sequence
38	150	5.1	280050	1	AL591975	AL591975 Listeria
39	150	5.1	349980	6	AX641665	AX641665 Sequence
40	147.4	5.1	1340	6	AX416059	AX416059 Sequence
41	147.4	5.1	189050	1	AL646077	AL646077 Ralstonia
42	142.6	4.9	300029	1	AE016952	AE016952 Enterococ
43	141	4.8	2684	1	ENBERSA	L23802 Enterococcu
44	133	4.6	756	6	AX414335	AX414335 Sequence
45	129.2	4.4	28206	1	BACDIA	L09228 Bacillus su

ALIGNMENTS

RESULT 1	A47369	2883 bp	DNA	linear	PAT 07-MAR-1997
LOCUS	A47369	Sequence 1 from Patent WO9531553.			
DEFINITION	A47369				
ACCESSION	A47369.1	GI:2301369			
VERSION					
KEYWORDS					
SOURCE	Neisseria polysacchara				
ORGANISM	Neisseria polysacchara				
	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;				
REFERENCE	1 (bases 1 to 2883)				
AUTHORS	Kossmann, J., Buettcher, V. and Welsh, T.				
TITLE	DNA SEQUENCES CODING FOR ENZYMES CAPABLE OF FACILITATING THE SYNTHESIS OF LINEAR alpha -1,4 GLUCANS IN PLANTS, FUNGI AND				

Pred. No. is the number of results predicted by chance to have a

MICROORGANISMS
 Patent: WO 9531553-A 1 23-NOV-1995;
 INST GENBIOLOGISCHE FORSCHUNG (DE)
 COMMENT
 Other publication DE 4447388 960627
 Other publication AU 2614195 951205.
 FEATURES
 Location/Qualifiers
 1..2883
 /organism="Neisseria polysacchara"
 /mol_type="genomic DNA"
 /db_xref="taxon:489"
 /clone="PNB2"
 /clone_lib="GENOMIC LIBRARY IN PLBUESCRIPTII SK"
 939..2783
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA03058.1"
 /db_xref="GI:2301370"
 /translation="MLTPTQVGLILQYKTRILDIYTPQAGIEKSDWRQFRRM
 DTFPKLNNELDSVYGNBHALPMLKLELAQWYSQNSLSKXDIDIARENNPWIL
 SNKQVGVYVDLFDAGLKLKDKIPYFELGLTYLHMLPLFKPSPGSKSDGAVSYT
 RDVNPALGTIGDLRRLVIAALHSHFRRRFYLPQRLQTRMAQRCAQGLPFDNFYIIF
 PRMRPDQYDRLRIEIPDHPGSPQLEDGRVMTFNSPQMDLNSPNVFAQWRA
 KCCLPTWALITSCVMWMLPLFGNKQQAATCAAHALLIPAFNAVMELIAPAVFPKSEA
 IVHPQVQYITGDRCOIGTNELQWALLNLTATREVNLLHQALTYRNLPERHTAVN
 YVSRHDDIGWTFADSDAIVLGSYDHRQFLNRFVNRFDGTFARGVFPQINPSTGDC
 RYSGTAAALVLAQDDPHAVDRIKLLYSIALSTGLLEPLIYLGVDEVTINDDDWCQAAI
 RATTAVGPSLQSPVATERSVDRSQIYQGLRHMIAVROSNPFRPDGRLVNTNTN
 NKHIIGYIATMRFMSVTSANIRKPLRIPICKPCPSRRTTSSVAKLSA"
 561 a 852 c 758 g 612 t
 BASE COUNT
 ORIGIN
 Query Match 90.6%; Score 2641; DB 6; Length 2883;
 Best Local Similarity 97.8%; Pred. No. 0;
 Matches 2852; Conservative 0; Mismatches 30; Indels 33; Gaps 16;
 QY 1 GAGTTTGGTTCCTCCGAAACGAAAGTGATGCTTGAGCCGAAACACCTGTCCGCGAAGCGG 60
 DB 1 GAGTTTGGTTCCTCCGAAACGAAAGTGATGCTTGAGCCGAAACACCTGTCCGCGAAGCGG 60
 QY 61 CTGACCGCCCTTTTGCCCAATCGACATCGTAAACAAATCGTTGGTGCAAGCTCTTTC 120
 DB 61 -----ACCGCTTTTGCCCAATCGACATCGTAAACAAATCGTTGGTGCAAGCTCTTTC 114
 QY 121 GCTTCAGCGGTGCGAAGAACGAAAGTCAGACGCTTTCGCGGTTCGCGCATCACCGCA 180
 DB 115 GCTTCAGCGGTGCGAAGAACGAAAGTCAGACGCTTTCGCGGTTCGCGCATCACCGCA 173
 QY 181 ATTTTCAGATGTCCGCGCGAGTCTCCATCTGTTTCAGACGCGATACGATTTCTTCT 240
 DB 174 ATTTTCAGATGTCCGCGCGAGTCTCCATCTGTTTCAGACGCGATACGATTTCTTCT 233
 QY 241 TCGCGCGGTGCGGTGAACCTATGATTCAGACGAGCGCGCGATGCGGTTTTTTTGA 300
 DB 234 TCGCGCGGTGCGGTGAACCTATGATTCAGACGAGCGCGGTGCGGTTTTTTTGA 290
 QY 301 GCATTCGCAACGCGCGCGCGGAGCGCGGTTTCGCGGAAAGAAAGCTCGATATCGATAATG 360
 DB 291 GCATTCGCAACGCGCGCGG-----AGCGGTTTCGCGGAAAGAAAGCTCGATATCGATAATG 344
 QY 361 TCGCGACGCGCGGTTCATTCAGCGAGTTCAGACGAGTTCAGAAATAATTCGTCGGAACAC 420
 DB 345 TCGCGACGCGCGGTTCATTCAGCGAGTTCAGACGAGTTCAGAAATAATTCGTCGGAACAC 404
 QY 421 GCGAACGAGCGCGCTTCGCGCATGCGCTCTGAAAGTAAAGCAAGCAGCGCGCTTTCGCGCAGC 480
 DB 405 GCGAACGAGCGCGCTTCGCGCATGCGCTCTGAAAGTAAAGCAAGCAGCGCGCTTTCGCGCAGC 464
 QY 481 GCGTCGCGGACGCTCTGCGGTGCGCGCAATCTTCGCGGATGCTCCCGCGCATTCGAAA 540
 DB 465 GCGTCGCGGACGCTCTGCGGTGCGCGCAATCTTCGCGGATGCTCCCGCGCATTCGAAA 524
 QY 541 AATTCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTTGGTTGGTCAAGTACGCGG 600

DB 525 AATTCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTTGAATTCGTCGAAGTACGCGG 584
 QY 601 GAAAGTACGCGCGCATTCGCGGCGACAAAGCGGACGCGGATTTTGGTGCCTCCGCTCCG 660
 DB 585 GAAAGTACGCGCGCATTCGCGGCGACAAAGCGGACGCGGATTTTGGTGCCTCCGCTCCG 644
 QY 661 ATAACGGTGTGTTTTCAGCGTCA-GGCTGGTGTGCATGCGCGTGTTCGCGGTGAAGAA 719
 DB 645 ATAACGGTGTGTTTTCAGCGTCAAGCGTTCAGCGTGTTCGCGGTGAAGAA 704
 QY 720 CGGTAAAGACCAATTTATAGCAAGGACAGGCGAATTTTCAGACGCGCATTTCTGTGCGG 779
 DB 705 CGGTAAAGACCAATTTATAGCAAGGACAGGCGAATTTTCAGACGCGCATTTCTGTGCGG 764
 QY 780 CCGGCTTGATATGAATCAAGCAGCATTCGCGATATCGGAATCGACATTTGGCAAGCCCT 839
 DB 765 CCGGCTTGATATGAATCAAGCAGCATTCGCGATATCGGAATCGACATTTGGCAAG-CCT 823
 QY 840 GTCTTTTCTAGTCAGTCGCGAGTTCCTTCAGTATGATTTGCAGACGACGCGCTTACCGCA 899
 DB 824 GTCTTTTCTAGTCAGTCGCGAGTTCCTTCAGTATGATTTGCAGACGACGCGCTTACCGCA 883
 QY 900 TTTCAGGATACGCGCGGACAGCCGCGGTCCGAAACTTCAGAAATCGGAGCAGGCATCATG 959
 DB 884 TTTCAGGATACGCGCGGACAGCCGCGGTCCGAAACTTCAGAAATCGGAGCAGGCATCATG 941
 QY 960 TTGACCCCCACGCGAGAGTGGTGTGATTTTACGATACCTCAAAAACAGCATCTTGAC 1019
 DB 942 TTGACCCCCACGCGAGAGTGGTGTGATTTTACGATACCTCAAAAACAGCATCTTGAC 1001
 QY 1020 ATCTACGCGCGGAACTAGCGCGCGCATTCGAAATTCGGAAGACTGCGCGGCGATTTTCG 1079
 DB 1002 ATCTACGCGCGGAACTAGCGCGCGCATTCGGAATTCGGAAGACTGCGCGGCGATTTTCG 1061
 QY 1080 CCGCGCATGATACGATTTCCCAAACTGATGAAAGAACTGCAAGCGGTGTAACGCAAC 1139
 DB 1062 CCGCGCATGATACGATTTCCCAAACTGATGAAAGAACTGCAAGCGGTGTAACGCAAC 1121
 QY 1140 AACGAAGCCCTGCTGCTATCTGGAATGCTGCGCGAGGCATGCAAGCTGATTTCC 1199
 DB 1122 AACGAAGCCCTGCTGCTATCTGGAATGCTGCGCGAGGCATGCAAGCTGATTTCC 1181
 QY 1200 CAACGCAACTCATCTTTAAAGATATCGATATCGCGCGGCGGAAACAAACCCGATTTGAT 1259
 DB 1182 CAACGCAACTCATCTTTAAAGATATCGATATCGCGCGGCGGAAACAAACCCGATTTGAT 1241
 QY 1260 TTGTCCAAACAAAGTCCGCGCGGTGCTACGTTGATTTGTTGCGCGGATTTGAAG 1319
 DB 1242 TTGTCCAAACAAAGTCCGCGCGGTGCTACGTTGATTTGTTGCGCGGATTTGAAG 1301
 QY 1320 GCTTGAAGATATAAATTCCTTATTTTCAAGAGCTTGGTTTGAATTTCTGCACTGATG 1379
 DB 1302 GCTTGAAGATATAAATTCCTTATTTTCAAGAGCTTGGTTTGAATTTCTGCACTGATG 1361
 QY 1380 CCGCTGTTTAAATGCCCTGAAAGGCGAAGACGCGCGCTATGCGTCAAGCTACCGC 1439
 DB 1362 CCGCTGTTTAAATGCCCTGAAAGGCGAAGACGCGCGCTATGCGTCAAGCTACCGC 1421
 QY 1440 GATGTCAATTCGCGCACTGCGGCAATAGCGCACTTTCGCGAAGTCAATTCGCTGCGTGCAC 1499
 DB 1422 GATGTCAATTCGCGCACTGCGGCAATAGCGCACTTTCGCGAAGTCAATTCGCTGCGTGCAC 1481
 QY 1500 GAAGCGCGCATTTCCGCGGTGCTGATTTTATCTTCAACCAACACCTTCAACGAAACGAA 1559
 DB 1482 GAATC--GCAATTCGCGGTGCTGATTTTATCTTCAACCAACACCTTCAACGAAACGAA 1539
 QY 1560 TGGGCGCAACCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1619
 DB 1540 T--GGGCGCAACCTG--GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1595
 QY 1620 GACCGCGGATGCGCGCAATACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1679

1596 GACCGCGGATGCCGACCAATACGACCGACCCCTGCGGAAATCTTCCCGACGACAC 1655
1680 CCGGGGGCTTCTCGCACTGGAAGACGACGCGCTGCTGTGAGACACCTTCAATTCCTTC 1739
1656 CCGGGGGCTTCTCGCACTGGAAGACGACGCGCTGCTGTGAGACACCTTCAATTCCTTC 1715
1740 CAATGGGACTTGAATTTACAGCAACCGTGGGTATTCGCGCAATCGCGGGCGAAATGCTG 1799
1716 CAATGGGACTTGAATTTACAGCAACCGTGGGTATTCGCGCAATCGCGGGCGAAATGCTG 1774
1800 TTCTTGGCAACTTGGGCGCTTGAATTCCTGCTGATGAGATGCGGTTCCTTTATTTGGAAA 1859
1775 TTCTTGGCAACTTGGGCGCTTGAATTCCTGCTGATGAGATGCGGTTCCTTTATTTGGAAA 1834
1860 CAATGGGCAAGCTGCGAAACCTGCGCGAGCGACGCTCATCGCGCGCTTCAAT 1919
1835 CAATGGGCAAGCTGCGAAACCTGCGCGAGCGACGCTCATCGCGCGCTTCAAT 1892
1920 GCGGTATCGGTATTCGCGCGCGCGCGGTGTTCTTCAATCGGAAGCCATGTCACGCC 1979
1893 GCGGTATCGGTATTCGCGCGCGCGCGGTGTTCTTCAATCGGAAGCCATGTCACGCC 1952
1980 GACCAAGCTGCTCAATACATCGGCGAGGAGCAATCGCAATCGGTACAGCCCTGCA 2039
1953 GACCAAGCTGCTCAATACATCGGCGAGGAGCAATCGCAATCGGTACAGCCCTGCA 2012
2040 ATGGCATTTCTTGGAAACACCTTGGCCAGCGCGAGTCAACCTGCTCATCAGCGCTG 2099
2013 ATGGCATTTCTTGGAAACACCTTGGCCAGCGCGAGTCAACCTGCTCATCAGCGCTG 2072
2100 ACTACCGCCACACCTGCGCGAGCATACCGCTGCTCACTAGCTGCGGAGCAAGAC 2159
2073 ACTACCGCCACACCTGCGCGAGCATACCGCTGCTCACTAGCTGCGGAGCAAGAC 2132
2160 GACATCGGCTGACGTTTCCGATGAAGACGCGGATATCTGGGATAGCGCTACGAC 2219
2133 GACATCGGCTGACGTTTCCGATGAAGACGCGGATATCTGGGATAGCGCTACGAC 2192
2220 CACCGCAATTCCTCAACCGCTTCTTGGTCAACCGCTTCAACCGCTTCAACCGCTT 2279
2193 CACCGCAATTCCTCAACCGCTTCTTGGTCAACCGCTTCAACCGCTTCAACCGCTT 2252
2280 GTACGGTTCATACACCAACCAAGCAGCGGCTGCGGCTGCTGAGTGTACGCGCGCA 2339
2253 GTACGGTTCATACACCAACCAAGCAGCGGCTGCGGCTGCTGAGTGTACGCGCGCA 2312
2340 TTGGTTCGGCTTGGCGCAAGACGATCCCGCGCTTACCGCATCAAACTCTTGTACAGC 2399
2313 TTGGTTCGGCTTGGCGCAAGACGATCCCGCGCTTACCGCATCAAACTCTTGTACAGC 2372
2400 ATTGCTTTGAGTACCGGCGCTGCGGCTGATTTACTTAGGCGAGCGAGTGGTACGCTC 2459
2373 ATTGCTTTGAGTACCGGCGCTGCGGCTGATTTACTTAGGCGAGCGAGTGGTACGCTC 2432
2460 AATGACGAGCTGCTGCGAAGACACCAATAGAGCGAGCGACGCGCTTGGCGGACCGT 2519
2433 AATGACGAGCTGCTGCGGCGAG--CAGCATTAAGAGCGAGCGAGCGCTTGGCG--CAGCGT 2490
2520 CCGGCTCAACGAAACCTTGTACGCGCAACGCAACGATCGGTGACGCGCGGCA 2579
2491 CCGGCTCAACGAAACCTTGTACGCGCAACGCAACGATCGGTGACGCGCGGCA 2549
2580 ATCTACGGGCTTGGCGCATATGATTCGCTGCGCGCAAGCAATCCGCGCTTTCAGCGC 2639
2550 ATCTACGGGCTTGGCGCATATGATTCGCTGCGCGCAAGCAATCCGCGCTTTCAGCGC 2609
2640 GSCAGGCTGGTATCTCAACCAACCAAGCAACATCATCGGTACATCCGCAACAT 2699
2610 GSCAGGCTGGTATCTCAACCAACCAAGCAACATCATCGGTACAT--CGCAACAT 2668
2700 GCGCTTTGGCATTCGGTACTTCAGGATATTCGCAACCGTTACCGGATACCCCTG 2759
2669 GCGCTTTGGCATTCGGTACTTCAGGATATTCGCAACCGTTACCGGATACCCCTG 2728

QY 2760 CAAGCCATGCGCTTCAAGGCGCAGCAGCTCATCGCTGCGCAAACTGTGAGCTGAATCAG 2819
DB 2729 CAAGCCATGCGCTTCAAGGCGCAGCAGCTCATCGCTGCGCAAACTGTGAGCTGAATCAG 2788
QY 2820 GATTGACGCTTCAGCCCTATCAGCTCATGCTGCTGCGAAATCGCTGACGACGCTTCCC 2879
DB 2789 GATTGACGCTTCAGCCCTATCAGCTCATGCTGCTGCGAAATCGCTGACGACGCTTCCC 2848
QY 2880 AAATGCGCTTCAAGCGCTTTCAGACGCGCATTTGCG 2914
DB 2849 AAATGCGCTTCAAGCGCTTTCAGACGCGCATTTGCG 2883

RESULT 2

A47785 2883 bp DNA linear PAT 07-MAR-1997
LOCUS Sequence 1 from Patent WO9533838.
DEFINITION A47785

ACCESSION A47785

VERSION A47785.1 GI:2301671

KEYWORDS

SOURCE Neisseria polysaccharea

ORGANISM

Neisseria polysaccharea

Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

Neisseriaceae; Neisseria.

1 (bases 1 to 2883)

Kosman, J., Buettcher, V. and Welsh, T.

MICROORGANISMS PERMITTING THE INTRACELLULAR POLYHYDROXY ALKANOATE

SYNTHESIS WITH SIMULTANEOUS EXTRACELLULAR POLYSACCHARIDE SYNTHESIS

AND PROCESSES FOR PRODUCING THE SAME

PATENT: WO 9533838-A 1 14-DEC-1995;

INST GENBIOLOGISCHE FORSCHUNG (DF)

Other publication AU 2787895 960104

Other publication DE 4420223 950504.

FEATURES

source

1. 2883

/organism="Neisseria polysaccharea"

/mol_type="genomic DNA"

/db_xref="taxon:489"

/clone="PNB2"

/clone_lib="GENOMIC LIBRARY IN PBLUESCRIPTII SK"

939. 2783

/note="unnamed protein product"

/codon_start=1

/translation="MLPTQGVGLIOLYKTRILDIYTPQRIAGIEKSEDMRQFRRM

DTHPKLMNLDVYGNNEALLPMLLAQWQSYQNSLSKIDIDTARENPDML

SNKQGVYVDLFAGLKGLKIPYQELGLTFLHMLPLFKPEKSDGGYAVSTY

RDVNPALGTIDLRVIAALHESHRFRFYLPHLQRTMAQRCAGDPLDFNYIF

PDREPDQYDRLREIIPDHPGFSQLEDGRWYWTNSQWDIYNSPVAQWRA

KCCSLPTVALTSCVMRLPLFGNKWQDAKCAHALIRAFNAVMRIAPAVFKEEA

IVHPDVQYVIGQDEQIGYNPLQALLNLTATREVNLLHQALTYIRNLPEHTAWN

YVRSHDIDGTFADEDAAYLIGSYDHRQLFRFVNRFDGTFGVGFQYNPGDC

RVSCTAALVGLAQDPAVDRIKLYLSIALSTGLPLIYLGDEVTGLNDWDQAAI

RAFTAVQPSALQSPVRAERSVDRSQIYQGLRHMAVRSNPRFQDGLVTFNTN

NKHLIGIATWRFNHSVTSANIRKPLPCKPCPSRRTTSVAKLSA"

BASE COUNT 661 a 852 c 758 g 612 t

ORIGIN

Query Match 90.6%; Score 2641; DB 6; Length 2883;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 2852; Conservative 0; Mismatches 30; Indels 33; Gaps 16;

QY 1 GAGTTTTCGTTCCCGAACCGAACCTGATGCTTGGACCGCAACCTGTCGCGAAGCGG 60

DB 1 GAGTTTTCGTTCCCGAACCGAACCTGATGCTTGGACCGCAACCTGTCGCGAAGCGG 60

QY 61 CTGACCGCGCCCTTTTGGCCCATCGACATCGTAACAATCGGTTGTTGGTGGCAAGCTCTTTC 120

DB 61 -----ACCGCTTTTGGCCCATCGACATCGTAACAATCGGTTGTTGGTGGCAAGCTCTTTC 114

QY 121 GCTTTGAGCGTGGCAGAAAGCAAGTCAGCAGTCTTCCGCGCTTTGGGCGATCAACGCA 180
DB 115 GCTTTGAGCGTGGCAGAAAGCAAGTCAGCAGTCTTCCG - GCGTTTGGCGATCAACGCA 173
QY 181 ATTTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGGGATACGATTTCTTCT 240
DB 174 ATTTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGGGATACGATTTCTTCT 233
QY 241 TCGGGGGGCTGGGTTGAAACTCATGATTTGCAAGCAGGGCGCGATGCGGTTTGTGA 300
DB 234 TCGGGGGGCTGGGTTGAAACTCATGATTTGCAAGCAG - - - GCGATGCGGTTTGTGA 290
QY 301 GATGCGCCACGCGCGCGCGAGTCTTCCGCGGAAAGTTCGATATCGGTAATG 360
DB 291 GATGCGCCACGCGCGCG - - - - - AGCGGTTTCCGCGGAAAGTTCGATATCGGTAATG 344
QY 361 TCGGGCAGCGGCTTTCAATTCAGCGAGTCGAGCGATTCAAAATAAATATATGTCGCGAAGC 420
DB 345 TCGGGCAGCGGCTTTCAATTCAGCGAGTCGAGCGATTCAAAATAAATATATGTCGCGAAGC 404
QY 421 GGGAAAGAGCGGCTTTCGCGCATCGCGTCTGAAGTAAACAGCAGCGGCTTGTGCGGCGAGC 480
DB 405 GGGAAAGAGCGGCTTTCGCGCATCGCGTCTGAAGTAAACAGCAGCGGCTTGTGCGGCGAGC 464
QY 481 GCGTCGCGAGCGGCTTTCGCGCATCGCGTCTGAAGTAAACAGCAGCGGCTTGTGCGGCGAGC 540
DB 465 GCGTCGCGAGCGGCTTTCGCGCATCGCGTCTGAAGTAAACAGCAGCGGCTTGTGCGGCGAGC 524
QY 541 AAATCGCGCGGAGCTTCGAGCATATCGAAGGCGCGGTTTTCGATTTGTCGAGTACGCGG 600
DB 525 AAATCGCGCGGAGCTTCGAGCATATCGAAGGCGCGGTTTTCGATTTGTCGAGTACGCGG 584
QY 601 GAAAGTACGCGCGGCTTCGCGCGAGCAAGCGGCACTTCGTCGCGGCTTCG 660
DB 585 GAAAGTACGCGCGGCTTCGCGCGAGCAAGCGGCACTTCGTCGCGGCTTCG 644
QY 661 ATAAACGCTGTTTTCGCGTCA - GCGTCTGTCGATGCGGTTTTCGATTTGTCGAGTACGCGG 719
DB 645 ATAAACGCTGTTTTCGCGTCAAGGCTGCTGTCGATGCGGCTTTCGATTTGTCGAGTACGCGG 704
QY 720 CGGTAAAGACGCAATATAGCAAGGCAAGCGGCACTTCGAGCGGCTTCGTCGCGG 779
DB 705 CGGTAAAGACGCAATATAGCAAGGCAAGCGGCACTTCGAGCGGCTTCGTCGCGG 764
QY 780 CCGGCTTGATATGAATCAAGAGAGCTTCGCAATATCGGAATGAGAGTTCGCGCAAGCGCT 839
DB 765 CCGGCTTGATATGAATCAAGAGAGCTTCGCAATATCGGAATGAGAGTTCGCGCAAGCGCT 823
QY 840 GTCTTTTCTAGTCAGTCGCGAGTCTTCGAGTATGATTCAGACAGCGGCTTCACAGGCA 899
DB 824 GTCTTTTCTAGTCAGTCGCGAGTCTTCGAGTATGATTCAGACAGCGGCTTCACAGGCA 883
QY 900 TTTGCAGGATACGCGCGCAGACGCGGTCGGAACCTTCAGAACTCGGAGCAGGCAATCATG 959
DB 884 TTTGCAGGATACGCGCGCAGACGCG - - - GTCGGAACCTTCAGAACTCGGAGCAGGCAATCATG 941
QY 960 TTTGACCCCAAGCAGCAAGTTCGTTGATTTTACAGTACCTCAAAACAGCACTTCGAGC 1019
DB 942 TTTGACCCCAAGCAGCAAGTTCGTTGATTTTACAGTACCTCAAAACAGCACTTCGAGC 1001
QY 1020 ATCTACAGCCGAAACAGCGCGCGGCTTCGAAATTCGAAAGCTTCGCGGCGAGTTTTCG 1079
DB 1002 ATCTACAGCCGAAACAGCGCGCGGCTTCGAAATTCGAAAGCTTCGCGGCGAGTTTTCG 1061
QY 1080 CCGCGCATGATACGCAATTTCCCAACTGATGAGAGCACTTCAGAGGCTTCAGCGGCAAC 1139
DB 1062 CCGCGCATGATACGCAATTTCCCAACTGATGAGAGCACTTCAGAGGCTTCAGCGGCAAC 1121
QY 1140 AACGAAGCCCTGCTGCTATGCTGGAATGCTGCTGCGCAGGCACTTCGCAAGCTATTTCC 1199
DB 1122 AACGAAGCCCTGCTGCTATGCTGGAATGCTGCTGCGCAGGCACTTCGCAAGCTATTTCC 1181
QY 1200 CAAACGCACTCATCTCTTAAAGATATCGATATCGCGCGCGGCAACACCCCGATTTGGATT 1259

DB 1182 CAAACGCACTCATCTCTTAAAGATATCGATATCGCGCGGCAACACCCCGATTTGGATT 1241
QY 1260 TTTGCCAAACAAAGTCGCGCGGTGCTACGTTGATTTGTTGCGGCGATTTGAAG 1319
DB 1242 TTTGCCAAACAAAGTCGCGCGGTGCTACGTTGATTTGTTGCGGCGATTTGAAG 1301
QY 1320 GCGTTGAAAGTAAATTTCTTATTTTCAGAGCTTGGTTTGAATTTATCTGCACTGATG 1379
DB 1302 GCGTTGAAAGTAAATTTCTTATTTTCAGAGCTTGGTTTGAATTTATCTGCACTGATG 1361
QY 1380 CCGCTGTTTAAATGCCCTGAAAGCGACGCGCGCTATGCGGTTCAGAGCTACCGC 1439
DB 1362 CCGCTGTTTAAATGCCCTGAAAGCGACGCGCGCTATGCGGTTCAGAGCTACCGC 1421
QY 1440 GATGTCATTCGCGCACTGGGCGACATAGGCGACTTGGCGAGTCAATTCGTCGCGTGCAC 1499
DB 1422 GATGTCATTCGCGCACTGGGCGACATAGGCGACTTGGCGAGTCAATTCGTCGCGTGCAC 1481
QY 1500 GAAACCGGCACTTCGCGCGTCTGATTTTATCTTCAACCAACCTTCCCAACGCAACGAA 1559
DB 1482 GAATC - - - GCAATTCGCGCGTCTGATTTTATCTTCAACCAACCTTCCCAACGCAACGAA 1539
QY 1560 TGGGCGCAACGCTGCGCGCGCGCGCGCTTTCGACATTTCTACTATTTTCCCG 1619
DB 1540 T - GCGCGCAACGCTG - - - GCGCGCGCGCGCGCTTTCGACATTTTCTACTATTTTCCCG 1595
QY 1620 GACCGCGGATTCGCGCACTTCGCGCACTTCGCGCGGCTTTCGCGGAAATCTTCCCGCGACGAGC 1679
DB 1596 GACCGCGGATTCGCGCACTTCGCGCACTTCGCGCGGCTTTCGCGGAAATCTTCCCGCGACGAGC 1655
QY 1680 CCGCGCGGCTTCTCGCAACTGGAAGACGAGCGCTGGTGTGAGACGCTTCAATTCCTTC 1739
DB 1656 CCGCGCGGCTTCTCGCAACTGGAAGACGAGCGCTGGTGTGAGACGCTTCAATTCCTTC 1715
QY 1740 CAATGGGACTTGAATTAACAGCAACCGCTGGGATTT - CCGCAATGGCGGCGGAAATGCTG 1799
DB 1716 CAATGGGACTTGAATTAACAGCAACCGCTGGGATTT - CCGCAATGGCGGCGGAAATGCTG 1774
QY 1800 TTTCTTGCACCTTTGGGCTTGAATCTGCGTATGATGCGGTTCCTTTATTTGGAAA 1859
DB 1775 TTTCTTGCACCTTTGGGCTTGAATCTGCGTATGATGCGGTTCCTTTATTTGGAAA 1834
QY 1860 CAATGGGAGCAGCTCGAAGCTTCGCGCAGCGACGCGCTCATCTCCGCGGTTCAAT 1919
DB 1835 CAATGGGAGCAGCTCGAAGCTTCGCGCAGCGACGCGCTCATCTCCGCGGTTCAAT 1892
QY 1920 GCGGTTATGCGTATTTGCGCGCGCGCTGTTCTTCAAAATCCGAAGCCATCGTCCACCCC 1979
DB 1893 GCGGTTATGCGTATTTGCGCGCGCGCTGTTCTTCAAAATCCGAAGCCATCGTCCACCCC 1952
QY 1980 GACCAAGTCTCAATATATCGCGCAGGAGCAATGCGCAATCGGTTACACCCCTGCA 2039
DB 1953 GACCAAGTCTCAATATATCGCGCAGGAGCAATGCGCAATCGGTTACACCCCTGCA 2012
QY 2040 ATGCAATTTGTTGGAACACACCTTTGCGCAGCGCGAAGTCAACCTGCTCCATCAGCGGCTG 2099
DB 2013 ATGCAATTTGTTGGAACACACCTTTGCGCAGCGCGAAGTCAACCTGCTCCATCAGCGGCTG 2072
QY 2100 ACCTACCGCAGCAACCTTCCCGAGCATACGCGCTGGTCACTACGTCGCGAGCAGCAGC 2159
DB 2073 ACCTACCGCAGCAACCTTCCCGAGCATACGCGCTGGTCACTACGTCGCGAGCAGCAGC 2132
QY 2160 GACATCGGCTGGAATTTGCGGATGAAGACGCGCGCATATCTGGGCAATAAGCGGCTACGAC 2219
DB 2133 GACATCGGCTGGAATTTGCGGATGAAGACGCGCGCATATCTGGGCAATAAGCGGCTACGAC 2192
QY 2220 CACCGCAATTCCTCAACCGCTTCTTGTCAACCGTTTGAAGCGGAGCTTCGCTGCTGGC 2279
DB 2193 CACCGCAATTCCTCAACCGCTTCTTGTCAACCGTTTGAAGCGGAGCTTCGCTGCTGGC 2252
QY 2280 GTACCGTTCCAAATCAACCCAGCAGCGGCTGCGGCTGTCAGTGGTACAGCGCGGCA 2339


```
Db 2253 GTACCGTTCACATACACCAAGCAGCGGACTGCGGTGTCAGTGTACAGCCCGGCGA 2312
Qy 2340 TTGGTGGCTTGGCGCAGACGATCCCGACGGTGTGACCGCATCAAACTCTGTACAGC 2399
Db 2313 TTGGTGGCTTGGCGCAGACGATCCCGACGGTGTGACCGCATCAAACTCTGTACAGC 2372
Qy 2400 ATTGCTTTGAGTACCGCGGCTGCGCGTGTATTAACCTAGGCGAGAGTGGTACGCTC 2459
Db 2373 ATTGCTTTGAGTACCGCGGCTGCGCGTGTATTAACCTAGGCGAGAGTGGTACGCTC 2432
Qy 2460 AATGACGACGCTGGTGCAGACGACAAATAGACGACGACGCGTGTGGCGCACCGT 2519
Db 2433 AATGACGACGCTGGTGCACAG-CAGCAATAAGACGACGACGCGTGTGGC-CACCGT 2490
Qy 2520 CCGCGCTACACGAGCCCTGTACCGGCAACGCAAGCATCGTTCGACCGCGCGGCAA 2579
Db 2491 CCGCGCTACACGAGCCCTGTACCGGCAACGCAAGCATCGTTCGACCGCGCGGCAA 2549
Qy 2580 ATCTATCAGGGCTTGGCGCATATGATTGCGTCCGCGCAAGCAATCCGCGCTTCGACGCG 2639
Db 2550 ATCTATCAGGGCTTGGCGCATATGATTGCGTCCGCGCAAGCAATCCGCGCTTCGACGCG 2609
Qy 2640 GCGAGCTGTTACATTAACACCAACACACAGCAGCATCGCTACATCCGCAACAT 2699
Db 2610 GCGAGCTGTTACATTAACACCAACACACAGCAGCATCGCTACATCCGCAACAT 2668
Qy 2700 GCGCTTTTGGCATTCGGTAACTTCAGCGAATATCCGCAACCGTTACCGCGCATACCCCTG 2759
Db 2669 GCGCTTTTGGCATTCGGTAACTTCAGCGAATATCCGCAACCGTTACCGCGCATACCCCTG 2728
Qy 2760 CAAGCATGCTTCAAGCGCAGACCTCATCGGTGGCAAACTGTACGCTGAATCAG 2819
Db 2729 CAAGCATGCTTCAAGCGCAGACCTCATCGGTGGCAAACTGTACGCTGAATCAG 2788
Qy 2820 GATTTGAGCTTCAGCGCTCATCAGCTCATGCTGCGTGAATCGCTGACGACGCTTCC 2879
Db 2789 GATTTGAGCTTCAGCGCTCATCAGCTCATGCTGCGTGAATCGCTGACGACGCTTCC 2848
Qy 2880 AAATCCCGTCTGACCGTTCAGCGCATTTTGG 2914
Db 2849 AAATCCCGTCTGACCGTTCAGCGCATTTTGG 2883

RESULT 3
INFO011781
LOCUS Neisseria polysaccharea amylosucrase gene.
DEFINITION Neisseria polysaccharea amylosucrase gene.
ACCESSION AJ011781
VERSION AJ011781.1 GI:4049491
KEYWORDS amylosucrase; amylosucrase gene.
SOURCE Neisseria polysaccharea
ORGANISM Neisseria polysaccharea
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE 1
AUTHORS De Montalk, G.P., Renaud-Simeon, M., Willemot, R.M., Planchot, V. and
Monsan, P.
TITLE Sequence analysis of the gene encoding amylosucrase from Neisseria
polysaccharea and characterization of the recombinant enzyme
JOURNAL J. Bacteriol. 181 (2), 375-381 (1999)
MEDLINE 99102197
PUBMED 9882648
REFERENCE 2 (bases 1 to 2115)
AUTHORS Potocki de Montalk, G.
TITLE Direct submission
JOURNAL Submitted (02-OCT-1998) Potocki de Montalk G., U.M.R. C.N.R.S.
5504, L.A. I.N.R.A. D.G.B.A., I.N.S.A., Centre de biogenierie
Gilbert Durand, Complexe scientifique de Rangueil, Toulouse cedex,
31077, FRANCE
FEATURES
source
1..2115
/organism="Neisseria polysaccharea"
/mol_type="genomic DNA"
```

```
/strain="ATCC 43768"
/db_xref="taxon:489"
136..2046
/gene="amylosucrase"
<136..2046
/gene="amylosucrase"
/EC_number="2.4.1.4"
/codon_start=1
/transl_table=11
/product="amylosucrase"
/protein_id="CAA09772.1"
/db_xref="GI:4107260"
/transl_table="MLPTQOVLILQVLTLDIIVTPQORAGIEKSEDRWQFSRRL
DTHPKLWELDSVYGNELLLPMLLELQAQWOSYDORNSLXIDIDARNNPDWIL
SNKQVGVGYVDLPAQDLKGLDKIPLYOELGLTYLHMLPKPCKSGDGYAVSY
ADNVPALGCTIGDLREVIALHEAGISAVDFIENHTSNEHEWAORCAAGDPLDFNYI
IPDRMPDQYDRTLREIFPDHPGSPQVEDQWVHTFNSFOWDLNYSNPNWFRAM
AGEKFLANGLVDILRMDAVATWKGMTSCENLPQAHALIRANVRIRAAAPVFFK
SEATVHPDQVQYTGDEQCTGYNPLQALLWNTLATREVNLLHQAITYRENLEHTA
WNTVRGDDIGTFAEDDAAYLQISGVDRHQFLNRFFVNRFDGSGFARGVFOYNPST
GDCRVSTAAALVGLAODDPAVDRIKLKYSIALSTGCLPLIYLGDEVNPTLNDWMSQ
DSKSDSRWHRPRYNEALYAQRDPSSTAAGQIYQGLRHMIARQSNPRFDGRELVT
FNTNNKILIGTIRNNALLAFNFSFPQVTAHTLQAMPFKAHDLIGKTVSLNQDLT
IQPVQVWMLETA"
BASE COUNT 493 a 661 c 529 g 432 t
ORIGIN
Query Match 71.8%; Score 2093; DB 1; Length 2115;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2093; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 822 AGACTTGGCAACAAGCCCTGCTTTCTTAGTCAGTCGCGAGTCTTGCAGTAGTAGTTGCAC 881
Db 1 AGACTTGGCAACAAGCCCTGCTTTCTTAGTCAGTCGCGAGTCTTGCAGTAGTAGTTGCAC 60
Qy 882 GACGACCCCTACACGCGCATTTTCAGGATACCGCGGAGACGCGCGTGGGAACTTCAGA 941
Db 61 GACGACCCCTACACGCGCATTTTCAGGATACCGCGGAGACGCGCGTGGGAACTTCAGA 120
Qy 942 ATCGGAGCAGGCATCATGTTGACCCCGACGACGAGTCTGCTTTGATTTTACAGTACCTC 1001
Db 121 ATCGGAGCAGGCATCATGTTGACCCCGACGAGTCTGCTTTGATTTTACAGTACCTC 180
Qy 1002 AAAACAGCATCTTGGACATCTACGCGCGAAGCAGCGCGCGCATCGAAATCCGAA 1061
Db 181 AAAACAGCATCTTGGACATCTACGCGCGAAGCAGCGCGCGCATCGAAATCCGAA 240
Qy 1062 GACTGGCGGCGAGTTTTCGCGCGCATGATACGCAATTCGCCAAACTGATGAAAGACTC 1121
Db 241 GACTGGCGGCGAGTTTTCGCGCGCATGATACGCAATTCGCCAAACTGATGAAAGACTC 300
Qy 1122 GACAGGTGTACGCGCAACAGAGAGCCCTGCTGCTATGCTGGAATAGCTGCTGCGCGAG 1181
Db 301 GACAGGTGTACGCGCAACAGAGAGCCCTGCTGCTATGCTGGAATAGCTGCTGCGCGAG 360
Qy 1182 GCATGCGAAAGCTATTCGCAAGCAATCTCATCTCTTAAAGATATCGATATCGCGCGCA 1241
Db 361 GCATGCGAAAGCTATTCGCAAGCAATCTCATCTCTTAAAGATATCGATATCGCGCGCA 420
Qy 1242 AACCAACCCCGATTGCTTGTCCAAACAAACAGTCCGCGCGGTGTGCTAGCTGATTTG 1301
Db 421 AACCAACCCCGATTGCTTGTCCAAACAAACAGTCCGCGCGGTGTGCTAGCTGATTTG 480
Qy 1302 TTTGCGCGGCGATTTGAAGGCTTGAAAGATAAATTCCTTATTTTCAGAGCTTGGTTTG 1361
Db 481 TTTGCGCGGCGATTTGAAGGCTTGAAAGATAAATTCCTTATTTTCAGAGCTTGGTTTG 540
Qy 1362 ACTTATCTGACCTGTAGTCCGCTGTTTAAATGCCCTGAAGGCAAAAGCGAGCGCGCTAT 1421
Db 541 ACTTATCTGACCTGTAGTCCGCTGTTTAAATGCCCTGAAGGCAAAAGCGAGCGCGCTAT 600
Qy 1422 GCGGTGAGGAGTACCGGATGTCATTCGCGCACTGGCGCAATAGCGAGCTTGGCGAA 1481
```

Db 601 GCGGTGAGGAGTACCGCGATGTAATCGGCACTGGGCAATAGGCGACTTGGCGAA 660
Qy 1482 GTCAATGTCGCTGCAAGAGCGGCAATTCGCGCTGCTGCAATTTATTTCAACCAAC 1541
Db 661 GTCAATGTCGCTGCAAGAGCGGCAATTCGCGCTGCTGCAATTTATTTCAACCAAC 720
Qy 1542 ACCTCAAGAAACAGAAATGGGCGCAAGCTGCGCGCGCGGCGACCGCTTTTCGACAA 1601
Db 721 ACCTCAAGAAACAGAAATGGGCGCAAGCTGCGCGCGCGGCGACCGCTTTTCGACAA 780
Qy 1602 TTCTACTATATTTTCCCGACCGCGGATGCGCGCAATACGACCGACCGCTGGCGAA 1661
Db 781 TTCTACTATATTTTCCCGACCGCGGATGCGCGCAATACGACCGACCGCTGGCGAA 840
Qy 1662 ATCTTCCCGACGACGACCGCGGCTTCTCGCACTGCGCAAGCGCTGGGCTGG 1721
Db 841 ATCTTCCCGACGACGACCGCGGCTTCTCGCACTGCGCAAGCGCTGGGCTGG 900
Qy 1722 AGACCTTCAATTCCTTCCAAATGGGACTTGAATACAGCAACCGCTGGTATTCGCGCA 1781
Db 901 AGACCTTCAATTCCTTCCAAATGGGACTTGAATACAGCAACCGCTGGTATTCGCGCA 960
Qy 1782 ATGGCGGCGAATGCTGTTCTTCCAAATGGGACTTGAATACAGCAACCGCTGGTATTCGCGCA 1841
Db 961 ATGGCGGCGAATGCTGTTCTTCCAAATGGGACTTGAATACAGCAACCGCTGGTATTCGCGCA 1020
Qy 1842 GTTGCCTTTATTTGGAAACAAATGGGCAAGCTGCGAAACCTGCGCGAGCGCGCAAGCC 1901
Db 1021 GTTGCCTTTATTTGGAAACAAATGGGCAAGCTGCGAAACCTGCGCGAGCGCGCAAGCC 1080
Qy 1902 CTGATCGCGGCTTCAATGCGGTATGCGGTATGCGCGCGCGCGCGCTGCTTCAATGCG 1961
Db 1081 CTGATCGCGGCTTCAATGCGGTATGCGGTATGCGCGCGCGCGCGCTGCTTCAATGCG 1140
Qy 1962 GAGGCACTGCTCCACCGCGCAAGCTGCGCAATACATCGGCGAGCAAGCAATCCAAATC 2021
Db 1141 GAGGCACTGCTCCACCGCGCAAGCTGCGCAATACATCGGCGAGCAAGCAATCCAAATC 1200
Qy 2022 GGTACAAACCTTGGAAATGGCAATGTTGGAACACCTTGGCGAGCGCGCAAGTCAAC 2081
Db 1201 GGTACAAACCTTGGAAATGGCAATGTTGGAACACCTTGGCGAGCGCGCAAGTCAAC 1260
Qy 2082 CTGCTCATGAGCGCTGAGCTACCGCAACCTGCGCGAGCATACCGCTGGTCAAC 2141
Db 1261 CTGCTCATGAGCGCTGAGCTACCGCAACCTGCGCGAGCATACCGCTGGTCAAC 1320
Qy 2142 TAGCTCGCGACCGACGACGATCGCTGGAAGCTTTGCGGATGGAAGCGCGCATATCTG 2201
Db 1321 TAGCTCGCGACCGACGACGATCGCTGGAAGCTTTGCGGATGGAAGCGCGCATATCTG 1380
Qy 2202 GGCAATAGCGCTACGACGACCGCAATTCCTCAACCGCTTCTGCTCAACCGCTTTCGAC 2261
Db 1381 GGCAATAGCGCTACGACGACCGCAATTCCTCAACCGCTTCTGCTCAACCGCTTTCGAC 1440
Qy 2262 GGCAGCTTCTGCTGCGGTACCGTTTCCAAATACAAACCAAGCAGCGCGACTGCGGTGTC 2321
Db 1441 GGCAGCTTCTGCTGCGGTACCGTTTCCAAATACAAACCAAGCAGCGCGACTGCGGTGTC 1500
Qy 2322 AGTGGTACAGCGCGGCAATGGTTCGCTTGGCGCAAGCAATCCCGCGCGCTTTCGCGC 2381
Db 1501 AGTGGTACAGCGCGGCAATGGTTCGCTTGGCGCAAGCAATCCCGCGCGCTTTCGCGC 1560
Qy 2382 ATCAAACTCTTTGACAGATTTCTTTCAGTACCGCGGCTTCTGCGCTGATTTACTAGGC 2441
Db 1561 ATCAAACTCTTTGACAGATTTCTTTCAGTACCGCGGCTTCTGCGCTGATTTACTAGGC 1620
Qy 2442 GACCAAGTGGGTACGCTCAATGACGACGATGTTGCGCAAGCAGCAATACAGCGAGC 2501
Db 1621 GACCAAGTGGGTACGCTCAATGACGACGATGTTGCGCAAGCAGCAATACAGCGAGC 1680
Qy 2502 AGCGGTTGGCGCGACCGCTGCAACGACCGCTGACGCGCAAGCAGCAATACAGCGAGC 2561
Db 1681 AGCGGTTGGCGCGACCGCTGCAACGACCGCTGACGCGCAAGCAGCAATACAGCGAGC 1740

Qy 2562 TCGACCGGAGCGGCGCAATCTATCAGGCTTGGCGCATATGATTCGCTGCGCAAGC 2621
Db 1741 TCGACCGGAGCGGCGCAATCTATCAGGCTTGGCGCATATGATTCGCTGCGCAAGC 1800
Qy 2622 AATCGCGCTTTCGAGCGGCGGCGGTGTTTCAATTCATTCACCAACCAAGCACATATC 2681
Db 1801 AATCGCGCTTTCGAGCGGCGGCGGTGTTTCAATTCATTCACCAACCAAGCACATATC 1860
Qy 2682 GGTACATCCGCAACAAATGGCTTTTGGCATTCGGTAACTTCAGGAAATATCCGCAAC 2741
Db 1861 GGTACATCCGCAACAAATGGCTTTTGGCATTCGGTAACTTCAGGAAATATCCGCAAC 1920
Qy 2742 GTTACCGGCGATACCTTCCAAATGGCGCTTCAAGCGCGACGACCTCATCGTGGCAAA 2801
Db 1921 GTTACCGGCGATACCTTCCAAATGGCGCTTCAAGCGCGACGACCTCATCGTGGCAAA 1980
Qy 2802 ACTGTACGCTGAATTCAGGATTTGAGCTTTCAGCGCTTATCAGGCTATGCTGGCTCGAAATC 2861
Db 1981 ACTGTACGCTGAATTCAGGATTTGAGCTTTCAGCGCTTATCAGGCTATGCTGGCTCGAAATC 2040
Qy 2862 GCTTGAAGCAGCTTCCCAATGGCGCTTCAAGCGCTTTCAGCGCGCATTTTGG 2914
Db 2041 GCTTGAAGCAGCTTCCCAATGGCGCTTTCAGCGCGCATTTTGG 2093

RESULT 4

AX107013 LOCUS 1939 bp DNA linear PAT 30-APR-2001
Sequence 4 from Patent WO0125449.

AX107013 ACCESSION

AX107013.1 GI:13922547

VERSION

synthetic construct

KEYWORDS

artificial sequences.

ORGANISM

Bengs, H., Polakowski, T., Held, A. and Gallert, K. C.

Novel immobilizable amylosucrase, the use thereof and method for

producing poly(1,4-alpha-glucan)

Patent: WO 0125449-A 4 12-APR-2001;

AXIVA GmbH (DE)

FEATURES

Location/Qualifiers

1..1939

/organism="synthetic construct"

/mol_type="genomic DNA"

/db_xref="taxon:32630"

/note="PCR-Amplikon, kodierend fur amylosucrase (Amsu)"

BASE COUNT 453 a 605 c 482 g 395 t 4 others

ORIGIN

Query Match 65.9%; Score 1921.4; DB 6; Length 1939;
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;
Matches 1922; Conservative 0;

Qy 955 TCATGTTGACCCCGACGCGCAAGTCTGTTGATTTACAGTACCTCAAAACACGCTCT 1014

Db 10 TCATGTTGACCCCGACGCGCAAGTCTGTTGATTTACAGTACCTCAAAACACGCTCT 69

Qy 1015 TGGACATCTACGCGCGGACGCGCGCGGATCGAAATTCGAAATTCGAAATTCGAAATTC 1074

Db 70 TGGACATCTACGCGCGGACGCGCGCGGATCGAAATTCGAAATTCGAAATTCGAAATTC 129

Qy 1075 TTTTCGCGCGCATGGAATGATGCTTTCCCAAACTGATGAACGAATTCGACAGCGTACG 1134

Db 130 TTTTCGCGCGCATGGAATGATGCTTTCCCAAACTGATGAACGAATTCGACAGCGTACG 189

Qy 1135 GCAACAAAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194

Db 190 GCAACAAAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 249

Qy 1195 ATTCCCAACGCAACTCATCTTAAAGATATCATATCGCGCGCGGCAAAACAAACCCGAT 1254

250 ATTCCCAACGCACTCTCTTAAAGATATCGATATCGCGCGGAAACAAACCCCGATT 309
1255 GGATTTGTCACAAACAAAGTCGGCGCGTGTCTAGTGTGATTTGTTTGGCGCGGATT 1314
310 GGATTTGTCACAAACAAAGTCGGCGCGTGTCTAGTGTGATTTGTTTGGCGCGGATT 369
1315 TGAAGGCGCTTGAAGAGATAAATTCCTTATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 1374
370 TGAAGGCGCTTGAAGAGATAAATTCCTTATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 429
1375 TGATGCGCGCTTTTAAATGCGCTGAGAGCAAAAGAGCGCGGCTATGCGGTCAAGAGCT 1434
430 TGATGCGCGCTTTTAAATGCGCTGAGAGCAAAAGAGCGCGGCTATGCGGTCAAGAGCT 489
1435 ACCGCGATGTCAATCGGCACTGGGCAAAATAGGCGACTTGGCGGAGTCAATTCCTCGC 1494
490 ACCGCGATGTCAATCGGCACTGGGCAAAATAGGCGACTTGGCGGAGTCAATTCCTCGC 549
1495 TGCACGAAGCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 1554
550 TGCACGAAGCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 609
1555 ACGAATGGCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 1614
610 ACGAATGGCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 669
1615 TCCCGGACCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 1674
670 TCCCGGACCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 729
1675 AGCAGCGCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 1734
730 AGCAGCGCGCGGATTTCCGCGCGTGTGATTTTCAAGAGCTTGGTTTGNCTTATCTGCACC 789
1735 CTTTCCCAATGGGATTTGGAATACAGCAACCGTGGGTATTCGCGGCAATGGCGGCGGAAA 1794
790 CTTTCCCAATGGGATTTGGAATACAGCAACCGTGGGTATTCGCGGCAATGGCGGCGGAAA 849
1795 TGCTGTTCTTGCACAACTTGGCGGCTTGACATCTCTGCGGTATGGATGCGGTTGCGCTTTATTT 1854
850 TGCTGTTCTTGCACAACTTGGCGGCTTGACATCTCTGCGGTATGGATGCGGTTGCGCTTTATTT 909
1855 GGAAACAAATGGGACAGCTTGCAGAAACCTGCGCGGCGGACCGCTCATTCGCGCGCT 1914
910 GGAAACAAATGGGACAGCTTGCAGAAACCTGCGCGGCGGACCGCTCATTCGCGCGCT 969
1915 TCAATGCGGTTATGCGTATTCGCGCGGCGGCGGCTTCTTCAAAATCGGAGCCATCGTCC 1974
970 TCAATGCGGTTATGCGTATTCGCGCGGCGGCGGCTTCTTCAAAATCGGAGCCATCGTCC 1029
1975 ACCCGGACCAAGTCGTCACATACGCGGAGGAGGATGCAATCGGTTGAGTACACCCCG 2034
1030 ACCCGGACCAAGTCGTCACATACGCGGAGGAGGATGCAATCGGTTGAGTACACCCCG 1089
2035 TGCAATGGGATTTGTTGGGACACCCCTTGCACGCGGAGTCAACCTGCTCCATCAGG 2094
1090 TGCAATGGGATTTGTTGGGACACCCCTTGCACGCGGAGTCAACCTGCTCCATCAGG 1149
2095 CGCTGACCTACCGCACAACTGCGCGGAGGATACCGCTGGGTCACTAGTTCGCGAGCC 2154
1150 CGCTGACCTACCGCACAACTGCGCGGAGGATACCGCTGGGTCACTAGTTCGCGAGCC 1209
2155 ACAGACGACATCGGCTGAGAGTTTGGGATGAAGACGCGGATATCTGGGATAGAGCGCT 2214
1210 ACAGACGACATCGGCTGAGAGTTTGGGATGAAGACGCGGATATCTGGGATAGAGCGCT 1269
2215 ACAGACGACATCGGCTGAGAGTTTGGGATGAAGACGCGGATATCTGGGATAGAGCGCT 2274
1270 ACAGACGACATCGGCTGAGAGTTTGGGATGAAGACGCGGATATCTGGGATAGAGCGCT 1329
2275 GTGCGGTACCGTTCATACACCAAGCAGCGGAGTGGCGGTGAGTGTACAGCGGATACAGCG 2334
1330 GTGCGGTACCGTTCATACACCAAGCAGCGGAGTGGCGGTGAGTGTACAGCGGATACAGCG 1389

2335 CGGCAATGCTGCGCTTGGCGCGGAGCGATCCCGACGCGGTTGACCGCATCAAACTCTTGT 2394
1390 CGGCAATGCTGCGCTTGGCGCGGAGCGATCCCGACGCGGTTGACCGCATCAAACTCTTGT 1449
2395 ACAGCATGCTGCTTTGAGTACCGGCGGTCTGCGGCTGATTTTACCTAGGCGAGCAAGTGGGTA 2454
1450 ACAGCATGCTGCTTTGAGTACCGGCGGTCTGCGGCTGATTTTACCTAGGCGAGCAAGTGGGTA 1509
2455 CGCTCAATGACGAGCTGCTGCGGAGGAGCGATTAAGAGCGAGCGAGCGGTTGGGCGC 2514
1510 CGCTCAATGACGAGCTGCTGCGGAGGAGCGATTAAGAGCGAGCGAGCGGTTGGGCGC 1569
2515 ACCTGTCGCGCTACAAACGAGCGCTTGTACGCGCAACGCAACGATCCGTCGACCGGAGCGG 2574
1570 ACCTGTCGCGCTACAAACGAGCGCTTGTACGCGCAACGCAACGATCCGTCGACCGGAGCGG 1629
2575 GCGCAATCTATCAGGCTTGGCGCATATGATTTGCGGTCGCGCAAGCAATCCGCGCTTGC 2634
1630 GCGCAATCTATCAGGCTTGGCGCATATGATTTGCGGTCGCGCAAGCAATCCGCGCTTGC 1689
2635 ACCGCGCGGCTGCTTACATTCACCAACCAAGCAACATCATCGGCTACATCGGCA 2694
1690 ACCGCGCGGCTGCTTACATTCACCAACCAAGCAACATCATCGGCTACATCGGCA 1749
2695 ACAATGCGCTTTTGGCATTCGGTAACTTCAAGCGAATATTCGCAAAACGTTTACCGGCGATA 2754
1750 ACAATGCGCTTTTGGCATTCGGTAACTTCAAGCGAATATTCGCAAAACGTTTACCGGCGATA 1809
2755 CCTGCAAGCCATGCGCTTCAAGCGGCGAGCGCTCATCGTGGCGCAAAACGTTTACGCGCTGA 2814
1810 CCTGCAAGCCATGCGCTTCAAGCGGCGAGCGCTCATCGTGGCGCAAAACGTTTACGCGCTGA 1869
2815 ATCAGATTTGACGCTTACGCGCTTACGCGCTTACGCGCTCATGCGTTCGAAATCGCTGACGCGC 2874
1870 ATCAGATTTGACGCTTACGCGCTTACGCGCTTACGCGCTCATGCGTTCGAAATCGCTGACGCGC 1929
2875 TTC 2877
1930 CTC 1932

RESULT 5
AX107015
LOCUS AX107015 5878 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 6 from Patent WO0125449.
ACCESSION AX107015
VERSION AX107015.1 GI:13922549
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Bengt, H., Polakowski, T., Heid, A. and Gallert, K.C.
TITLE Novel immobilizable amylase, the use thereof and method for
producing poly(1,4-alpha-glucan)
JOURNAL Patent: WO 0125449-A 6 12-APR-2001;
Axiva GmbH (DE)
FEATURES
source Location/Qualifiers
1. .6878
/organism="genomic DNA"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="Beschreibung der künftlichen
Sequenz: Expressionsvektor pGEX-4T-1-Amsu"
BASE COUNT 1673 a 1800 c 1766 g 1639 t
ORIGIN

Query Match 65.9%; Score 1921.4; DB 6; Length 6878;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1922; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 955 TCATGTTGACCCCGACGAGCAAGTGGTTGATTTTACAGTACCTCAAAACAGCGATCT 1014

943 TCATGTTGACCCCGACGAGTGGTTGATTTTACAGTACCTCAAAACAGCATCT 1002
1015 TGGACATCTACAGCCCGGAGAGCGCGGCGCATCGAATAATTCGAAGACTGGCGGAGT 1074
1003 TGGACATCTACAGCCCGGAGAGCGCGGCGCATCGAATAATTCGAAGACTGGCGGAGT 1062
1075 TTTTCGGCGCGGATGAGTACGATTTCCCAAACTGATGAAACGAACTCGACACGCTGTACG 1134
1063 TTTTCGGCGCGGATGAGTACGATTTCCCAAACTGATGAAACGAACTCGACACGCTGTACG 1122
1135 GCAACAAAGAGCCCTGCTGCTATGCTGGAAATGCTGCTGGGAGGAGCATGGCAAGCT 1194
1123 GCAACAAAGAGCCCTGCTGCTATGCTGGAAATGCTGCTGGGAGGAGCATGGCAAGCT 1182
1195 ATTCCCAACGCACTCATCTTAAGAGATATGATATCGCGCGGAGAAACAAACCCGAT 1254
1183 ATTCCCAACGCACTCATCTTAAGAGATATGATATCGCGCGGAGAAACAAACCCGAT 1242
1255 GGAATTTGTCGAACAAAGTGGCGGCTGTGCTACGTTGATTTGTTGTCGGCGGAT 1314
1243 GGAATTTGTCGAACAAAGTGGCGGCTGTGCTACGTTGATTTGTTGTCGGCGGAT 1302
1315 TGAAGGCTTGAAGATTAATTCCTATTTTCAAGAGCTTGGTTGATTTCTATCTGCACC 1374
1303 TGAAGGCTTGAAGATTAATTCCTATTTTCAAGAGCTTGGTTGATTTCTATCTGCACC 1362
1375 TGAATGCGGCTGTTTAAATGCCCTGAAGGCAAAAGACGACGCGGCTATGCGGTGACAGCT 1434
1363 TGAATGCGGCTGTTTAAATGCCCTGAAGGCAAAAGACGACGCGGCTATGCGGTGACAGCT 1422
1435 ACCGGAATGTCATTCGGGACCTGGGACAAATAGGCGACTTGGCGGAAGTCATGCTGGCG 1494
1423 ACCGGAATGTCATTCGGGACCTGGGACAAATAGGCGACTTGGCGGAAGTCATGCTGGCG 1482
1495 TGCACGAAGCGCGCATTTCCGCGCTGCTCGATTTTATCTTCAACACACACTCCAAACGAC 1554
1483 TGCACGAAGCGCGCATTTCCGCGCTGCTCGATTTTATCTTCAACACACACTCCAAACGAC 1542
1555 ACGAATGGGCGCAACGCTGCGCGCGCGGACCGCGCTTTTGGACAAATTTCTACTATATT 1614
1543 ACGAATGGGCGCAACGCTGCGCGCGCGGACCGCGCTTTTGGACAAATTTCTACTATATT 1602
1615 TCCCGGACCGCGGATTCGCGACCAATACGACGCGACCTGCGCGAAATCTTCCCGGAC 1674
1603 TCCCGGACCGCGGATTCGCGACCAATACGACGCGACCTGCGCGAAATCTTCCCGGAC 1662
1675 AGCACCGCGCGGCTTCTCGCAACTGGAAGACGACGCTGGGTGTGGAAGACCTTCAATT 1734
1663 AGCACCGCGCGGCTTCTCGCAACTGGAAGACGACGCTGGGTGTGGAAGACCTTCAATT 1722
1735 CCTTCCAAATGGGACTTGAATTAAGCAACCGGTGGGTATTCGCGCAATGGGCGGCAAA 1794
1723 CCTTCCAAATGGGACTTGAATTAAGCAACCGGTGGGTATTCGCGCAATGGGCGGCAAA 1782
1795 TGCCTGTTCTTTCGCAACTTGGGCGGTTGACATCTCTGCGTATGATGATGCGGTTGCTTTATT 1854
1783 TGCCTGTTCTTTCGCAACTTGGGCGGTTGACATCTCTGCGTATGATGATGCGGTTGCTTTATT 1842
1855 GGAACAAATGGGGAACGACTGCGAAGTCTGGCGAGGCGGACGCTGCTCATCGCGCGT 1914
1843 GGAACAAATGGGGAACGACTGCGAAGTCTGGCGAGGCGGACGCTGCTCATCGCGCGT 1902
1915 TCAATGCGGTTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAAAATCGGAAGCCATCGTC 1974
1903 TCAATGCGGTTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAAAATCGGAAGCCATCGTC 1962
1975 ACCCGGACCAAGTGTCTCAATATACATCGGGGAGGAGCAATGCGAAATCGGGTACAGCCGC 2034
1963 ACCCGGACCAAGTGTCTCAATATACATCGGGGAGGAGCAATGCGAAATCGGGTACAGCCGC 2022
2035 TGCAAATGGCAATGTTGTGGAACACCCCTGCGCGGAGTCAACCTGCTCCATCAGG 2094

2023 TGCAATGGCAATGTTGTGGAACACCTTGTCCACGCGGAGTCAAACTGCTGCTCCATCAGG 2082
2095 CGCTGACCTACCGCCACCACTGCTGCCGAGCATACGCTTGGGTCAAATCTAGCTCCGCAAGC 2154
2083 CGCTGACCTACCGCCACCACTGCTGCCGAGCATACGCTTGGGTCAAATCTAGCTCCGCAAGC 2142
2155 ACGACGACATCGCTTGGACGCTTTCGCAATGAAGACGCGGCAATATCTGGGCAATAAGCGGCT 2214
2143 ACGACGACATCGCTTGGACGCTTTCGCAATGAAGACGCGGCAATATCTGGGCAATAAGCGGCT 2202
2215 ACGACCAACCGCAATTCCTCAACCGCTTCTTCGTCACACCGCTTTCGACGCGAGCTTCGCTC 2274
2203 ACGACCAACCGCAATTCCTCAACCGCTTCTTCGTCACACCGCTTTCGACGCGAGCTTCGCTC 2262
2275 GTGGCGTACCGTTTCCAAATACCAACCAAGCACAGCGACTGCGGTGTCAAGTGGTACAGCG 2334
2263 GTGGCGTACCGTTTCCAAATACCAACCAAGCACAGCGACTGCGGTGTCAAGTGGTACAGCG 2322
2335 CGGCATTTGCTGCGCTTGGCGCAAGCGATCCCGCGCTTTCGACGCGATCAAACTCTTGT 2394
2323 CGGCATTTGCTGCGCTTGGCGCAAGCGATCCCGCGCTTTCGACGCGATCAAACTCTTGT 2382
2395 ACAGCATTTGCTTGAATACCGGCGCTCTGCGCTGATTTTACCTAGGCGACGAAGTGGTA 2454
2383 ACAGCATTTGCTTGAATACCGGCGCTCTGCGCTGATTTTACCTAGGCGACGAAGTGGTA 2442
2455 CGCTCAATGACGACGACTGCTGCGAAGACGAAATAGAGCGACGACGCGCTTGGCGC 2514
2443 CGCTCAATGACGACGACTGCTGCGAAGACGAAATAGAGCGACGACGCGCTTGGCGC 2502
2515 ACCGTCGCGCTTACCAACGAAGCCCTGTACGCGCAACGCAACGATCCGTCGACCGCAGCG 2574
2503 ACCGTCGCGCTTACCAACGAAGCCCTGTACGCGCAACGCAACGATCCGTCGACCGCAGCG 2562
2575 GCGAAATCTACAGGCTTGGCGCAATGATTTGCGCTGCGCAAGCAATCGCGCTTCG 2634
2563 GCGAAATCTACAGGCTTGGCGCAATGATTTGCGCTGCGCAAGCAATCGCGCTTCG 2622
2635 ACGCGGACGAGGCTGTTACATTTCAACCAACCAAGCAACATCATCGCTACATCCGCA 2694
2623 ACGCGGACGAGGCTGTTACATTTCAACCAACCAAGCAACATCATCGCTACATCCGCA 2682
2695 ACAATGCGCTTTTGGCAATCGGTAACTTCAAGCAATATCGCAAAACCGTTACCGGCGATA 2754
2683 ACAATGCGCTTTTGGCAATCGGTAACTTCAAGCAATATCGCAAAACCGTTACCGGCGATA 2742
2755 CCTCGAAGCATGCTTCAAGCGCGACGCTCATCGGTGGCAAACTGTACGCTGA 2814
2743 CCTCGAAGCATGCTTCAAGCGCGACGCTCATCGGTGGCAAACTGTACGCTGA 2802
2815 ATCAGGATTTGACGCTTCAGCCCTATCAGGTCATGCTGCTGAAATCGCTTCACGACGC 2874
2803 ATCAGGATTTGACGCTTCAGCCCTATCAGGTCATGCTGCTGAAATCGCTTCACGACGC 2862
2875 TTC 2877
2863 CTC 2865

RESULT 6
AX107014
LOCUS
DEFINITION
Sequence 5 from Patent WO0125449.
ACCESSION
AX107014
VERSION
AX107014.1 GI:13922548
KEYWORDS
synthetic construct
synthetic construct
artificial sequences.
ORGANISM
Bengals, H., Polakowski, T., Held, A. and Gallert, K. C.
REFERENCE
1
Novel immobilizable amylase, the use thereof and method for
producing poly(1,4-alpha-glucan)

AX107014 1910 bp DNA linear PAT 30-APR-2001
Sequence 5 from Patent WO0125449.

JOURNAL Patent: WO 0125449-A 5 12-APR-2001;
Axiva GmbH (DE)
FEATURES Location/Qualifiers
source 1. 1910
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="PCR-Amplikon, kodierend fur verkurzte amylosucrase
(AmSu5)"
BASE COUNT 447 a 595 c 474 g 387 t 7 others
ORIGIN
Query Match 64.9%; Score 1891.4; DB 6; Length 1910;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1892; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 985 TGAATTTACAGTACTCAAAACACGACATCTTGGACATCTACACGCCGGAACAGCGCGCG 1044
Db 11 TGAATTTACAGTACTCAAAACACGACATCTTGGACATCTACACGCCGGAACAGCGCGCG 70
Qy 1045 GCATCGAAAAATCCGAGACTGGCGGCAATTTTCGGCGCGCATGGATACGCAATTTCCCCA 1104
Db 71 GCATCGAAAAATCCGAGACTGGCGGCAATTTTCGGCGCGCATGGATACGCAATTTCCCCA 130
Qy 1105 AACTGATGAACGAACTCGACAGCTGTACGGCAACAGAGCCCTGCTGCTATGCTGG 1164
Db 131 AACTGATGAACGAACTCGACAGCTGTACGGCAACAGAGCCCTGCTGCTATGCTGG 190
Qy 1165 AAATGCTGCTGGCGAGGATCGCAAGACTATTCACACGCAACTCATCTCTAAAGATA 1224
Db 191 AAATGCTGCTGGCGAGGATCGCAAGACTATTCACACGCAACTCATCTCTAAAGATA 250
Qy 1225 TCGATATCGCGCGGAAACAAACCCGATTTGGATTTTGTCCAAACAAAGTCGGCGCG 1284
Db 251 TCGATATCGCGCGGAAACAAACCCGATTTGGATTTTGTCCAAACAAAGTCGGCGCG 310
Qy 1285 TGTGCTAGCTTGATTTGTTTCCGCGGATTTGAAGGGCTTGAAGATAAAATTCCTTATT 1344
Db 311 TGTGCTAGCTTGATTTGTTTCCGCGGATTTGAAGGGCTTGAAGATAAAATTCCTTATT 370
Qy 1345 TTCAGAGCTTGGTTGACTATCTGACCTGATGCGCTGTTTAAATGCCCTGAAGCA 1404
Db 371 TTCAGAGCTTGGTTGACTATCTGACCTGATGCGCTGTTTAAATGCCCTGAAGCA 430
Qy 1405 AAAGCGAGCGGCTATCGGTCAGCAGTACCGGATGTCAATCCGGCACTGGGCAAA 1464
Db 431 AAAGCGAGCGGCTATCGGTCAGCAGTACCGGATGTCAATCCGGCACTGGGCAAA 490
Qy 1465 TAGCGACTTGGCGAAGTCAATTGCTGCTGTCACGAAGCGCGCATTTCCGCGTGTGTC 1524
Db 491 TAGCGACTTGGCGAAGTCAATTGCTGCTGTCACGAAGCGCGCATTTCCGCGTGTGTC 550
Qy 1525 ATTTTATCTTCAACACACCTCCACGACACGATGGCGCAACGCTGCGCGCGCGCG 1584
Db 551 ATTTTATCTTCAACACACCTCCACGACACGATGGCGCAACGCTGCGCGCGCGCG 610
Qy 1585 ACCCGCTTTTCGCAATTTCTATATATTTTCCCGACCGCGGATGCCGACCAATACG 1644
Db 611 ACCCGCTTTTCGCAATTTCTATATATTTTCCCGACCGCGGATGCCGACCAATACG 670
Qy 1645 ACCGACCGCTGGCGAATCTTCCCGACCGACACCGCGGCGCTTCTGCAACTGGAAG 1704
Db 671 ACCGACCGCTGGCGAATCTTCCCGACCGACACCGCGGCGGCTTCTGCAACTGGAAG 730
Qy 1705 ACGGACGCTGGGTGTGGACGACCTTCAATTTCTTCCCAATGGGACTTGAATACAGCAACC 1764
Db 731 ACGGACGCTGGGTGTGGACGACCTTCAATTTCTTCCCAATGGGACTTGAATACAGCAACC 790
Qy 1765 CGTGGGTATTCGCGGATGGCGGGGAAATCTGTTCTGTCGCAACTGGCGTTGACA 1824
Db 791 CGTGGGTATTCGCGGATGGCGGGGAAATCTGTTCTTCCCAACTTGGCGTTGACA 850
Qy 1825 TCCTCGTATGATGCGGTTGCTTTATTTTGAACAAATGGGGCAAGCTGCGAAAACC 1884

RESULT 7

851 TCCTGCTATGATGCGGTTGCCCTTTATTTGAAAACAAATGGGACAAAGCTGCGAAAACC 910
Qy 1885 TGCCGCGAGCGCACGCCCTCATCCGCGCGCTTCAATGCGGTTATGCTATTTGCGCGCGCG 1944
Db 911 TGCCGCGAGCGCACGCCCTCATCCGCGCGCTTCAATGCGGTTATGCTATTTGCGCGCGCG 970
Qy 1945 CCGTGTCTTCAAAATCCGAAGCCATCGTCCACCCCGACCAAGTGTCTCAATATATCGGGC 2004
Db 971 CCGTGTCTTCAAAATCCGAAGCCATCGTCCACCCCGACCAAGTGTCTCAATATATCGGGC 1030
Qy 2005 AGGACGAATGCCAAATCGGTTACAAACCCCTGCAAAATGGCATTTGTTGGAACACACCTTTC 2064
Db 1031 AGGACGAATGCCAAATCGGTTACAAACCCCTGCAAAATGGCATTTGTTGGAACACACCTTTC 1090
Qy 2065 CCACGCGCGAAGTCAACCTGCTCCATCAGCGCTGACCTACCGCCCAACACCTGCGCGAGC 2124
Db 1091 CCACGCGCGAAGTCAACCTGCTCCATCAGCGCTGACCTACCGCCCAACACCTGCGCGAGC 1150
Qy 2125 ATACCGCTGGTCACTAGTCCGAGCGACACGACGACATCGGCTGGACGCTTTGCCGATG 2184
Db 1151 ATACCGCTGGTCACTAGTCCGAGCGACACGACGACATCGGCTGGACGCTTTGCCGATG 1210
Qy 2185 AAGAAGCGGATATCTGGGCAATAGCGGCTACGACCAACCGCCCAATTCCTCAACCGCTTCT 2244
Db 1211 AAGAAGCGGATATCTGGGCAATAGCGGCTACGACCAACCGCCCAATTCCTCAACCGCTTCT 1270
Qy 2245 TCGTCAACCGTTTCGACGCGGAGCTTCTGCTGCTGGGTACCGTTCCAAATCAACCCCAAGCA 2304
Db 1271 TCGTCAACCGTTTCGACGCGGAGCTTCTGCTGCTGGGTACCGTTCCAAATCAACCCCAAGCA 1330
Qy 2305 CAGCGCACTCCGCTGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2364
Db 1331 CAGCGCACTCCGCTGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1390
Qy 2365 CCACGCGCTTGCACCGCATCAAACTCTGTGTACAGCATTTGCTTTGAGTACCGCGCTGCTG 2424
Db 1391 CCACGCGCTTGCACCGCATCAAACTCTGTGTACAGCATTTGCTTTGAGTACCGCGCTGCTG 1450
Qy 2425 CGCTGATTTACCTAGCGCGAAGTGGTGTACGCTCAATGACGACGACTGTGTGCAAGACA 2484
Db 1451 CGCTGATTTACCTAGCGCGAAGTGGTGTACGCTCAATGACGACGACTGTGTGCAAGACA 1510
Qy 2485 GCAATAGAGCGACGACGCGCTTGGCGCACCGTCCGCGCTACACGAAGCCCTGTACG 2544
Db 1511 GCAATAGAGCGACGACGCGCTTGGCGCACCGTCCGCGCTACACGAAGCCCTGTACG 1570
Qy 2545 GCGAACGCAAGATCCGTCGACCGCGACCGCGGCAATCTATCAGGGCTTGGCGCATATGA 2604
Db 1571 GCGAACGCAAGATCCGTCGACCGCGACCGCGGCAATCTATCAGGGCTTGGCGCATATGA 1630
Qy 2605 TTGCGGTCGCGCAAAAGCAATCCGCGCTTTCGACCGCGCGGCGAGGCTGTTTACATTTCAACACCA 2664
Db 1631 TTGCGGTCGCGCAAAAGCAATCCGCGCTTTCGACCGCGCGGCGAGGCTGTTTACATTTCAACACCA 1690
Qy 2665 ACAACAGACGACATCATGCGCTTACATCCGCAACAAATGCGCTTTTGGGATTCGTTACTTCA 2724
Db 1691 ACAACAGACGACATCATGCGCTTACATCCGCAACAAATGCGCTTTTGGGATTCGTTACTTCA 1750
Qy 2725 GCGAATATCCGCAAAACCGTTTACCGCGCATACCTTGAAGCCATCGCCCTTCAAGCGCGCAAG 2784
Db 1751 GCGAATATCCGCAAAACCGTTTACCGCGCATACCTTGAAGCCATCGCCCTTCAAGCGCGCAAG 1810
Qy 2785 ACCTCATCGGTGGCAAACTGTGTGAGCTGAATGAGATTTGACGCTTTCAGCGCTTATCAGG 2844
Db 1811 ACCTCATCGGTGGCAAACTGTGTGAGCTGAATGAGATTTGACGCTTTCAGCGCTTATCAGG 1870
Qy 2845 TCATGTGCTCGAATCGCTGTACGCGCTTC 2877
Db 1871 TCATGTGCTCGAATCGCTGTACGCGCTTC 1903

AX107016
LOCUS AX107016 6851 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 7 from Patent WO0125449.
ACCESSION AX107016
VERSION AX107016.1 GI:13922550
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Bengs, H., Polakowski, T., Held, A. and Gallert, K.C.
TITLE Novel immobilizable amylosucrase, the use thereof and method for
producing poly(1,4-alpha-glucan)
JOURNAL Patent: WO 0125449-A 7 12-APR-2001;
Axiva GmbH (DE)
FEATURES
Location/Qualifiers
1..6851
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="Beschreibung der kunstlichen
Sequenz: Expressionsvektor pGEX-4T-1-AmSu5"
BASE COUNT 1668 a 1791 c 1759 g 1633 t
ORIGIN
Query Match 64.9%; Score 1891.4; DB 6; Length 6851;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1892; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 985 TGATTTTACAGTACCTCAAAACAGCATCTTCGACATCTACACGCCCGCAAGAGCGCGC 1044
DB 946 TGATTTTACAGTACCTCAAAACAGCATCTTCGACATCTACACGCCCGCAAGAGCGCGC 1005
QY 1045 GCATCGAAATAATCCGAAGACTGCGGCGAGTTTCGCGCGCATGATAGCATTTCCCGCA 1104
DB 1006 GCATCGAAATAATCCGAAGACTGCGGCGAGTTTCGCGCGCATGATAGCATTTCCCGCA 1065
QY 1105 AACTGATGACGACGATCGACAGCGTGTACGCAAGCAAGCAAGCGCTGTGCTATGCTGG 1164
DB 1066 AACTGATGACGACGATCGACAGCGTGTACGCAAGCAAGCAAGCGCTGTGCTATGCTGG 1125
QY 1165 AAATGCTGTCGCGGAGGATGCGCAAGCTATTCGCAAGCGCACTCATCTTTAAAGATA 1224
DB 1126 AAATGCTGTCGCGGAGGATGCGCAAGCTATTCGCAAGCGCACTCATCTTTAAAGATA 1185
QY 1225 TCGATATCGGCGCGGAAACACCCCGATGGATTTGTTCGCAAGCAAGCTGCGGCGG 1284
DB 1186 TCGATATCGGCGCGGAAACACCCCGATGGATTTGTTCGCAAGCAAGCTGCGGCGG 1245
QY 1285 TGTGCTACGTTGATTTGTTTGGCGGCAATTGAGGCGCTTGAAGATAAAATTCCTTATT 1344
DB 1246 TGTGCTACGTTGATTTGTTTGGCGGCAATTGAGGCGCTTGAAGATAAAATTCCTTATT 1305
QY 1345 TTCAAGAGCTTGGTTTGACTTATCTGCACTGATGCGCTGTTTAAATGCCCTGAAGCA 1404
DB 1306 TTCAAGAGCTTGGTTTGACTTATCTGCACTGATGCGCTGTTTAAATGCCCTGAAGCA 1365
QY 1405 AAAGCGACCGCGCTATGCGGTGACAGCTACCGGATGCTCAATCCGCACTGGGCACAA 1464
DB 1366 AAAGCGACCGCGCTATGCGGTGACAGCTACCGGATGCTCAATCCGCACTGGGCACAA 1425
QY 1465 TAGGCGACTTGGCGGAAGTCATTTGCTGCGCTGCAAGACCGGCAATTCGCGCGTGTGG 1524
DB 1426 TAGGCGACTTGGCGGAAGTCATTTGCTGCGCTGCAAGACCGGCAATTCGCGCGTGTGG 1485
QY 1525 ATTTTATCTTACACACCTCCCAACGACGATGGCGGCAAGCTGCGCGCGCGG 1584
DB 1486 ATTTTATCTTACACACCTCCCAACGACGATGGCGGCAAGCTGCGCGCGCGG 1545
QY 1585 ACCCGCTTTGCAAAATTTCTACTATTTTCCCGACCGCGGATGCGCGCAATACG 1644
DB 1546 ACCCGCTTTGCAAAATTTCTACTATTTTCCCGACCGCGGATGCGCGCAATACG 1605

QY 1645 ACCGACCCCTGCGGAAATCTTCCCGACAGCACCCGCGCGGCTTCTCGCAACTGGAAG 1704
DB 1606 ACCGACCCCTGCGGAAATCTTCCCGACAGCACCCGCGCGGCTTCTCGCAACTGGAAG 1665
QY 1705 ACGGACGCTGGGTGTGAGACGCTTCAATTCCTTCCAAATGGGACTTGAATTTACGCAAC 1764
DB 1666 ACGGACGCTGGGTGTGAGACGCTTCAATTCCTTCCAAATGGGACTTGAATTTACGCAAC 1725
QY 1765 CGTGGTATTCGCGCAATGCGGCGGAAATGCTTCTTCCCAACTTGGGCGGTGACA 1824
DB 1726 CGTGGTATTCGCGCAATGCGGCGGAAATGCTTCTTCCCAACTTGGGCGGTGACA 1785
QY 1825 TCCTGCGGTATGAGATGCGGTGCTTTTATTGGAACAAATGGGACAAAGCTGCGCAAAAC 1884
DB 1786 TCCTGCGGTATGAGATGCGGTGCTTTTATTGGAACAAATGGGACAAAGCTGCGCAAAAC 1845
QY 1885 TGCGGACGCGCACGCTCATTCGCGGCTTCAATGCGGTATGCTGCTATTCGCGCGCGG 1944
DB 1846 TGCGGACGCGCACGCTCATTCGCGGCTTCAATGCGGTATGCTGCTATTCGCGCGCGG 1905
QY 1945 CCGTGTTCCTCAAAATCCGAAGCCATGCTCCACCCCGACCAAGTGTGCTCAATACATCGGGC 2004
DB 1906 CCGTGTTCCTCAAAATCCGAAGCCATGCTCCACCCCGACCAAGTGTGCTCAATACATCGGGC 1965
QY 2005 AGGAGGATGCGCAATGCTTACAAACCCCTGCAATGGGACTTGTGGAACACCCCTTG 2064
DB 1966 AGGAGGATGCGCAATGCTTACAAACCCCTGCAATGGGACTTGTGGAACACCCCTTG 2025
QY 2065 CCAAGCGGAAAGTCAACCTGCTCCATCAGGCGCTGACCTACCGCCCAACACCTGCGCGAGC 2124
DB 2026 CCAAGCGGAAAGTCAACCTGCTCCATCAGGCGCTGACCTACCGCCCAACACCTGCGCGAGC 2085
QY 2125 ATAGCGCTGGGTCAACTAGCTCGGACGACGACGACATCGGCTGGAGCGTTTCCGCGATG 2184
DB 2086 ATAGCGCTGGGTCAACTAGCTCGGACGACGACGACATCGGCTGGAGCGTTTCCGCGATG 2145
QY 2185 AAGACGCGGCAATATCTGGGCATAAGCGGCTACGACCAACCGCAATTCCTCAACCGTTCT 2244
DB 2146 AAGACGCGGCAATATCTGGGCATAAGCGGCTACGACCAACCGCAATTCCTCAACCGTTCT 2205
QY 2245 TCGTCAACCGTTTTCGACGCGAGCTTCGCTGCTGGCGGTACCGTTTCCATACGACCCAGCA 2304
DB 2206 TCGTCAACCGTTTTCGACGCGAGCTTCGCTGCTGGCGGTACCGTTTCCATACGACCCAGCA 2265
QY 2305 CAGGCGACTGCGGTGAGTGGTACAGCGCGGCAATTTGCTCGGCTTGGCGCAAGACGATC 2364
DB 2266 CAGGCGACTGCGGTGAGTGGTACAGCGCGGCAATTTGCTCGGCTTGGCGCAAGACGATC 2325
QY 2365 CCCACGCGTTGACCGCATCAAACTCTTTGTAAGCATTTGCTTGGAGTACCGCGCGTCTGC 2424
DB 2326 CCCACGCGTTGACCGCATCAAACTCTTTGTAAGCATTTGCTTGGAGTACCGCGCGTCTGC 2385
QY 2425 CGCTGATTTACCTAGGCGAGCAAGTGGGTACGCTCAATGACGACGCTGGTCCGCAAGACA 2484
DB 2386 CGCTGATTTACCTAGGCGAGCAAGTGGGTACGCTCAATGACGACGCTGGTCCGCAAGACA 2445
QY 2485 GCAATAAGAGGAGCAGCAGCGGTTGGGCGCAACCGTCCGCGCTCAACGAAAGCCCTGTACG 2544
DB 2446 GCAATAAGAGGAGCAGCAGCGGTTGGGCGCAACCGTCCGCGCTCAACGAAAGCCCTGTACG 2505
QY 2545 CGCAACGCAACGATCCGTCGACCGCGGCAAAATCTATCAGGCTTGGCGCATATGA 2604
DB 2506 CGCAACGCAACGATCCGTCGACCGCGGCAAAATCTATCAGGCTTGGCGCATATGA 2565
QY 2605 TTGCGCTCCGCAAGCAATCCGCGCTTCGACGCGGCGAGGCTGGTTTACATTCAACACCA 2664
DB 2566 TTGCGCTCCGCAAGCAATCCGCGCTTCGACGCGGCGAGGCTGGTTTACATTCAACACCA 2625
QY 2665 ACAACAGCAGCATCATCGGCTACATCGGCAACATGCGCTTTTGGCATTCGGTAACCTCA 2724
DB 2626 ACAACAGCAGCATCATCGGCTACATCGGCAACATGCGCTTTTGGCATTCGGTAACCTCA 2685
QY 2725 CGGATATTCGCAAAACGCTTACCGCGCATACCTTCGCAAGCAATGCGCTTCAAGCGCGCAG 2784


```

Db      2686 GCGAATATCCGCAAAACCGTTACCGGCGCATACCTTCGAAGCCATGCGCTTCAAGGCGCAGC 2745
QY      2785 ACCTCATCGTGGCGAAACCTGTCAGCCTGAATCAGGATTGACGCTTCAGCGCCTATCAGG 2844
Db      2746 ACCTCATCGTGGCGAAACCTGTCAGCCTGAATCAGGATTGACGCTTCAGCCCTATCAGG 2805
QY      2845 TCATGTGGCTCGAATATGCGCTGACGACGCTTC 2877
Db      2806 TCATGTGGCTCGAATATGCGCTGACGACGCTTC 2838

RESULT 8
AY099334      1967 bp      DNA      linear      BCT 13-JAN-2003
LOCUS      Neisseria meningitidis strain 93246 amylosucrase Ams (ams) gene,
DEFINITION      complete cds.
ACCESSION      AY099334
VERSION      AY099334.1 GI:27728141
KEYWORDS
SOURCE      Neisseria meningitidis
ORGANISM      Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE      1 (bases 1 to 1967)
AUTHORS      Zhu, P., Teang, R.S.W. and Tsai, C.M.
TITLE      Nonencapsulated Neisseria meningitidis Strain Produces Amylopectin
from Sucrose: Altering the Concept for Differentiation between N.
meningitidis and N. polysacchara
JOURNAL      J. Clin. Microbiol. 41 (1), 273-278 (2003)
PUBMED      12517860
REFERENCE      2 (bases 1 to 1967)
AUTHORS      Zhu, P.
TITLE      Direct Submission
JOURNAL      Submitted (23-APR-2002) LBP/DBPAP, CBER, 8800 Rockville Pike,
Bethesda, MD 20892, USA
FEATURES
source
    1..1967
        /organism="Neisseria meningitidis"
        /mol_type="genomic DNA"
        /strain="93246"
        /db_xref="taxon:487"
    21..1931
        /gene="ams"
    21..1931
        /gene="ams"
        /codon_start=1
        /transl_table=11
        /product="amylosucrase Ams"
        /protein_id="AAMS1152.1"
        /db_xref="GI:27728142"
        /translation="MLTPTQVGLILQYLKTRILDIYTPHORAGIEKSEDWRFQSRRL
        DTHPKLNEIDSVYGNNEALLPMLLELAQWOSYORSASLKNIDIERENNPDWIL
        SNKQVGVYVDLFDGLKLDKIHVEQLGLTYLHMLPLPKPEKSDGAYVSSY
        RDVNPALGTIGDREIVIALHREAGISAVVDIEFNHNSHEHWQRCAAGDPLFNPYY
        IFPDRMPQVDRTREIFPDHPGGFQSLSDRWVWTFNSFQNDLNSNFWFRAM
        AGEMLPLAGVDILRMDAVAFIKWQMTSCENLPQAHALIRAFNAVMRIAPAVFFK
        SEAIIVHPQVQVYIGQDEQCIQYNPQLQWALLNTLATREVNLLQALTYRNLPEHTA
        WUNYVRSHDDIGWTFADDAAYLGIISYDHRQFLNRPVNRPDGSEARGVPFOYNPT
        GDCRVSGTSAALQAQNDHVAVSEIKLLYSIALSTGLPLIYLGDVGTDPNDGWQ
        DSNKDDSRWAHRPYNALYAQRNDPSTAGQIYQGLREMIAYRQSNRPFDGGRLYT
        FNNKHLIGIRRNALLAFNFSYPTVTAHLQAQFPFAHDLIGKTVSLNQDLT
        LQPYQVMWLEIA"
BASE COUNT      463 a      526 c      490 g      388 t
ORIGIN
    Query Match      63.6%; Score 1853.4; DB 1; Length 1967;
    Best Local Similarity 96.4%; Pred. No. 0;
    Matches 1896; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY      937 TCAGAATCGAGCAGGCATCATGTTGACCCCGCAGCAGCAAGTCGGTTGATTTTACAGT 996
Db      1 TCAGAATCGAGCAGGCATCATGTTGACCCCGCAGCAGCAAGTCGGTTGATTTTACAGT 60

```

```

QY      997 ACCTCAAAACACGCAATCTTGACATCTACAGCCCGCAACAGCGCGCGGCAATCGAAAAT 1056
Db      61 ACCTCAAAACACGCAATCTTGACATCTACAGCCCGCAACAGCGCGCGGCAATCGAAAAT 120
QY      1057 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 1116
Db      121 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 180
QY      1117 AACTCGACAGCGTGTACGGCAACAAACGAAGCCCTGCTGCTATCTGTGAAATGCTGCTGG 1176
Db      181 AACTCGACAGCGTGTACGGCAACAAACGAAGCCCTGCTGCTATCTGTGAAATGCTGCTGG 240
QY      1177 CCGAAGACTGCGGCAAGCTATTTCACACGCAACTCATCTTAAAGATATCGATATCGCGC 1236
Db      241 CCGAAGACTGCGGCAAGCTATTTCACACGCAACTCATCTTAAAGATATCGATATCGCGC 300
QY      1237 CCGAAGACTGCGGCAAGCTATTTCACACGCAACTCATCTTAAAGATATCGATATCGCGC 1296
Db      301 CCGAAGACTGCGGCAAGCTATTTCACACGCAACTCATCTTAAAGATATCGATATCGCGC 360
QY      1297 ATTTGTTTGGCGGCGATTGAAGGGCTGAAAGATAAATAATTCCTTATTTTCAAGAGCTTG 1356
Db      361 ATTTGTTTGGCGGCGATTGAAGGGCTGAAAGATAAATAATTCCTTATTTTCAAGAGCTTG 420
QY      1357 GTTTGACTTATCTGACATGATGCGCTGTTTAAATGCCCTGAAGCAAAAGCGACGCGCG 1416
Db      421 GTTTGACTTATCTGACATGATGCGCTGTTTAAATGCCCTGAAGCAAAAGCGACGCGCG 480
QY      1417 GCTATGCGGTGACGAGCTACCGGATGTCAATCGCGCACTGGGCAATAGCGGACTTGC 1476
Db      481 GCTATGCGGTGACGAGCTACCGGATGTCAATCGCGCACTGGGCAATAGCGGACTTGC 540
QY      1477 GCGAAGCTATGCTGCGCTGACACGACGCGCATTTCCGCGCTGCTGATTTTATCTTCA 1536
Db      541 GCGAAGCTATGCTGCGCTGACACGACGCGCATTTCCGCGCTGCTGATTTTATCTTCA 600
QY      1537 ACCACACTCTCAACGAAACAGAAATGGCGGCAACGCTGCGCGCGCGCGGCAACCGCTTTTCG 1596
Db      601 ACCACACTCTCAACGAAACAGAAATGGCGGCAACGCTGCGCGCGCGCGGCAACCGCTTTTCG 660
QY      1597 ACAATTTCTACTATATTTTCCCGACCGCGGATGCGCGACCAATACAGACGCGACCTGC 1656
Db      661 ACAATTTCTACTATATTTTCCCGACCGCGGATGCGCGACCAATACAGACGCGACCTGC 720
QY      1657 GCGAAATCTTCCCGACGACGACCGCGCGGCTTCTCGCAACTGGAAGACGAGCGCTGGG 1716
Db      721 GCGAAATCTTCCCGACGACGACCGCGCGGCTTCTCGCAACTGGAAGACGAGCGCTGGG 780
QY      1717 TGTGACGACCTTCAATTTCTTCCATGGGACTTGAATTAAGCAACCCGTTGGTATTC 1776
Db      781 TGTGACGACCTTCAATTTCTTCCATGGGACTTGAATTAAGCAACCCGTTGGTATTC 840
QY      1777 GCGCAATGGCGGCGAAATGCTGTTCTTCCCACTTGGGCTTGCATCTCTGCTATGG 1836
Db      841 GCGCAATGGCGGCGAAATGCTGTTCTTCCCACTTGGGCTTGCATCTCTGCTATGG 900
QY      1837 ATGCGGTGCTCTTTATTTGGAACAAATGGGACAACTGCGAAACCTGCGCGACGCGC 1896
Db      901 ATGCGGTGCTCTTTATTTGGAACAAATGGGACAACTGCGAAACCTGCGCGACGCGC 960
QY      1897 ACSCCTCATCTCCGCGCTTCAATGCGGTTATGGTATTTGGCGCGCGCGCTGTTCTTCA 1956
Db      961 ACSCCTCATCTCCGCGCTTCAATGCGGTTATGGTATTTGGCGCGCGCGCTGTTCTTCA 1020
QY      1957 AATCCGAAGCCATCGTCCACCGCCGACCAAGTCGTCCCAATATCATCGGCGAGGACGAATGCC 2016
Db      1021 AATCCGAAGCCATCGTCCACCGCCGACCAAGTCGTCCCAATATCATCGGCGAGGACGAATGCC 1080
QY      2017 AATCGGTTACACCCCTCCGAAATGCGATGTTGTGGAACACCCCTGCGCACCGCGCGAAG 2076
Db      1081 AATCGGTTACACCCCTCCGAAATGCGATGTTGTGGAACACCCCTGCGCACCGCGCGAAG 1140

```

2077 TCAACTGCTCCATCAGGCGCTGACCTACCGCCACAACTCCGAGCATACCGCTCGG 2136
Db |||||||
1141 TCAACTGCTCCATCAGGCGCTGACCTACCGCCACAACTCCGAGCATACCGCTCGG 1200
Qy |||||||
2137 TCAACTGCTCCATCAGGCGCTGACCTACCGCCACAACTCCGAGCATACCGCTCGG 2196
Db |||||||
1201 TCAACTGCTCCATCAGGCGCTGACCTACCGCCACAACTCCGAGCATACCGCTCGG 1260
Qy |||||||
2197 ATCTGGGATAGCGCTACGACCGCCAACTCTCGACCGCTCTCTCGTCAACCGTT 2256
Db |||||||
1261 ATCTGGGATAGCGCTACGACCGCCAACTCTCGACCGCTCTCTCGTCAACCGTT 1320
Qy |||||||
2257 TCGAGCGGAGCTTCGCTCGTGGCGTACCGTTCGACGACGACGACGACGACGACG 2316
Db |||||||
1321 TCGAGCGGAGCTTCGCTCGTGGCGTACCGTTCGACGACGACGACGACGACGACG 1380
Qy |||||||
2317 GTGTCACTGATAGCGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2376
Db |||||||
1381 GGTCACTGATAGCGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Qy |||||||
2377 ACCGATCAAACTCTGTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2436
Db |||||||
1441 GCGCATCAAACTCTGTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Qy |||||||
2437 TAGGCAAGGAGTGGTACGCTCAATGACGACGACGACGACGACGACGACGACG 2496
Db |||||||
1501 TCGGACGAGGAGTGGGACGCGGACGACGACGACGACGACGACGACGACGACG 1560
Qy |||||||
2497 ACGACGCGGTTGGGCGGACGCTGCGCGCTCAACGACGCTGACGCGGACGACG 2556
Db |||||||
1561 ACGACGCGGTTGGGCGGACGCTGCGCGCTCAACGACGCTGACGCGGACGACG 1620
Qy |||||||
2557 ATCGTTCGACGCGGCGGACGCTGCGCGCTCAACGACGCTGACGCGGACGACG 2616
Db |||||||
1621 ATCGTTCGACGCGGCGGACGCTGCGCGCTCAACGACGCTGACGCGGACGACG 1680
Qy |||||||
2617 AAAGCAATCGCGCTTCGACGCGGCGGACGCTGCGCGCTCAACGACGCTGACG 2676
Db |||||||
1681 AAAGCAATCGCGCTTCGACGCGGCGGACGCTGCGCGCTCAACGACGCTGACG 1740
Qy |||||||
2677 TCATCGGCTACATCCGCAACAACTGCGCTTTGCGCATTCGCTGATTCGCGAATATCG 2736
Db |||||||
1741 TCATCGGCTACATCCGCAACAACTGCGCTTTGCGCATTCGCTGATTCGCGAATATCG 1800
Qy |||||||
2737 AAACCGTTACCGGCGGATCGCTGACGCGGCGGACGCTGCGCGGCGGACGCTG 2796
Db |||||||
1801 AAACCGTTACCGGCGGATCGCTGACGCGGCGGACGCTGCGCGGCGGACGCTG 1860
Qy |||||||
2797 GCAAACTGTCAGCTGATTCAGGATTCGACGCTTCGACGCTTCGACGCTTCG 2856
Db |||||||
1861 GTHAACTGTCAGCTGATTCAGGATTCGACGCTTCGACGCTTCGACGCTTCG 1920
Qy |||||||
2857 AAATCGCTGACGCGCTTCGCAATGCGCTTCGCAATGCGCTTCGCAATGCGCT 2903
Db |||||||
1921 AAATCGCTGACGCGCTTCGCAATGCGCTTCGCAATGCGCTTCGCAATGCGCT 1967

RESULT 9
AY099335
LOCUS
DEFINITION
Neisseria polysacchara strain 85322 amylosucrase Ams (ams) gene,
complete cds.
ACCESSION
AY099335
VERSION
AY099335.1
KEYWORDS
GI:2728139
ORGANISM
Neisseria polysacchara
Neisseria polysacchara
Bacteria; Proteobacteria;
Neisseriaceae; Neisseria.
1 (bases 1 to 1967)
REFERENCE
Zhu, P., Tsang, R. S. W. and Tsai, C. M.
Nonencapsulated Neisseria meningitidis Strain Produces Amylopectin
from Sucrose: Altering the Concept for Differentiation between N.

meningitidis and N. polysacchara
J. Clin. Microbiol. 41 (1), 273-278 (2003)
12517860
2 (bases 1 to 1967)
Zhu, P.
Direct Submission
Submitted (23-APR-2002) LBP/DBPAP, CBER, 8800 Rockville Pike,
Bethesda, MD 20892, USA
FEATURES
Location/Qualifiers
1..1967
/organism="Neisseria polysacchara"
/mol_type="genomic DNA"
/strain="85322"
/db_xref="taxon:489"
21..1931
/gene="ams"
21..1931
/gene="ams"
/codon_start=1
/transl_table=1
/product="amylosucrase Ams"
/protein_id="AAMS1153.1"
/db_xref="GI:2728140"
/translation="MLPTQGVGLIQLHKLRIIDIVTPQORAGIEKSDWROFRRM
DTHEPLMNEIDSVYGNNEALLPMLEMLQAQMSYORSASLAKNIDIERNNPDWIL
SNKQVGVYVDLFAFGDLKGLKIPYQELGLTYLHMLPKPEKSDGAYVSSY
RDVNPALGTIDDLREVIALHEAGISAVVDIFNHTSNEHEWAORCAAGDLEFNFYI
IFDRMPDQYDRTREIFPDQHEGFSQLEDGEWVMTFNSFQMDLNSPWFVRA
AGEMLFLANGLDRLMDAVFIWKQGTSCENLPOAHALIRAFNAVRIAPAVFFK
SEAIVHPDQVQYIGQBCQGYNPQWALLNLTREVNLLHQALTIRNLEHTS
WNVYRSHDDIGMTFADSDAYLGLISYVDRHOFNRFVNRFDGSGFARVGFQNPNT
GDCVSTSAALAGLQNDHNAVSRILKLSIALSTGCLPLIYLIGDEVGTPTDQWAO
DGNKSDSRWHRPRINEALYQNRDSTAGQIYQGLRMIARQSNRPFQDGLVLT
FNTXKHIIIGYRNRLALFQNFSEYFQVTAHTLQAMPFAHDLIGKTVSLNQDLT
LQPYQVAMLEIA"
BASE COUNT 459 a 631 c 489 g 388 t
ORIGIN
Query Match 63.4%; Score 1847; DB 1; Length 1967;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 1892; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Qy 937 TCAGAAATCGGAGCAGGATCATGTTGACCCCGACGACGACGATCGTTTGATTTACAGT 996
Db 1 TCAGAAATCGGAGCAGGATCATGTTGACCCCGACGACGACGATCGTTTGATTTACAGC 60
Qy 997 ACCTCAAAACAGGATCTTGACATCTACACCGCGACGCGCGCGGATCGAAAT 1056
Db 61 ACCTCAAAACAGGATCTTGACATCTACACCGCGACGCGCGCGGATCGAAAT 120
Qy 1057 CCGAAGACTGCGCGGAGTTTTCGCGCGGATGATGATGATGATGATGATGATGATG 1116
Db 121 CCGAAGACTGCGCGGAGTTTTCGCGCGGATGATGATGATGATGATGATGATGATG 180
Qy 1117 AACTCGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1176
Db 181 AACTCGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
Qy 1177 CCGAGGATGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1236
Db 241 CCGAGGATGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Qy 1237 CCGAAACAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1296
Db 301 CCGAAACAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Qy 1297 ATTTGTTTTCGCGGAGTTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGG 1356
Db 361 ATTTGTTTTCGCGGAGTTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGG 420
Qy 1357 GTTTGACTTATCTGACCTGATGCGCTGTTTAAATGCCCTGAAAGGCAAAAGCGCGCG 1416
Db 421 GTTTGACTTATCTGACCTGATGCGCTGTTTAAATGCCCTGAAAGGCAAAAGCGCGCG 480

QY 1417 GCTATGCGGTGAGAGCTACCGCGATGTCAATCCGGGCACTGGGCAATAGGCGACTTGC 1476
DB GCTATGCGGTGAGAGCTACCGCGATGTCAATCCGGGCACTGGGCAATAGGCGACTTGC 540
QY 1477 GCGAAGTCAATGCTGGGTGCAAGAGCGGGCAATTCGCGGTGCTGCAATTTATCTTCA 1536
DB GCGAAGTCAATGCTGGGTGCAAGAGCGGGCAATTCGCGGTGCTGCAATTTATCTTCA 600
QY 1537 ACCACACCTCCAAAGCAAGAGTGGGCGCAAGCTGCGCGCGCGCGCGCGCGCTTTTCG 1596
DB ACCACACCTCCAAAGCAAGAGTGGGCGCAAGCTGCGCGCGCGCGCGCGCGCTTTTCG 660
QY 1597 ACAATTTCTACTATATTTTCCCGACCGCGGATGCCGCAATACAGCGACCGCTTGC 1656
DB ACAATTTCTACTATATTTTCCCGACCGCGGATGCCGCAATACAGCGACCGCTTGC 720
QY 1657 GCGAATTTCTCCGACGACCGCGGGGCTTCTCCGCACTGGAAGCAAGCGCTTGG 1716
DB GCGAATTTCTCCGACGACCGCGGGGCTTCTCCGCACTGGAAGCAAGCGCTTGG 780
QY 1717 TGTGACGACCTTCAATTTCTTCCAAATGGGACTTGAATTAAGCAACCGCTGGGTATTC 1776
DB TGTGACGACCTTCAATTTCTTCCAAATGGGACTTGAATTAAGCAACCGCTGGGTATTC 840
QY 1777 GCGAATGGCGGGCGAATGCTGTTCTTCCGCACTGCGGGTGGACATCTGCGTATGG 1836
DB GCGAATGGCGGGCGAATGCTGTTCTTCCGCACTGCGGGTGGACATCTGCGTATGG 900
QY 1837 ATGCGGTGCTTATTTGGAAACAAATGGGACCAAGCTGCGAACAACCTGCGGAGGGC 1896
DB ATGCGGTGCTTATTTGGAAACAAATGGGACCAAGCTGCGAACAACCTGCGGAGGGC 960
QY 1897 AGCGCTCAATCGCGGCTTCAATGCGGTATGCGTATTTGCGCGCGCGCGCGCTTCTTCA 1956
DB AGCGCTCAATCGCGGCTTCAATGCGGTATGCGTATTTGCGCGCGCGCGCGCTTCTTCA 1020
QY 1957 AATCGGAGCATGCTGCGACCGCGCAAGCTGCGAATACATACGCGGAGGACGAATGCC 2016
DB AATCGGAGCATGCTGCGACCGCGCAAGCTGCGAATACATACGCGGAGGACGAATGCC 1080
QY 2017 AATCGGTTACAAACCGCTGCAATGGCAATTTGTTGGAACCAACCTTCCCAACCGCGAAG 2076
DB AATCGGTTACAAACCGCTGCAATGGCAATTTGTTGGAACCAACCTTCCCAACCGCGAAG 1140
QY 2077 TCAACTGCTCCTCAGCGGCTGACCTACGCGCAACCTGCGCGGAGCATACCGGCTGGG 2136
DB TCAACTGCTCCTCAGCGGCTGACCTACGCGCAACCTGCGCGGAGCATACCGGCTGGG 1200
QY 2137 TCAACTAGCTCGGAGCGACACGACATCGGCTGACGCTTTGCGGATGAAGACGCGGAT 2196
DB TCAACTAGCTCGGAGCGACACGACATCGGCTGACGCTTTGCGGATGAAGACGCGGAT 1260
QY 2197 ATCTGGGATAGCGGCTAGCAACCGCGCAATTCCTCAACCGCTTCTTCTGCTCAACCGTT 2256
DB ATCTGGGATAGCGGCTAGCAACCGCGCAATTCCTCAACCGCTTCTTCTGCTCAACCGTT 1320
QY 2257 TCGACGGGAGCTTCTGCTGCGGCTACCGTTCATACCAACCGGACGACGCGGCTGCC 2316
DB TCGACGGGAGCTTCTGCTGCGGCTACCGTTCATACCAACCGGACGACGCGGCTGCC 1380
QY 2317 GTGTGAGTGTACAGCGCGGATTTGTTGCGGCTTGGCGCAAGACGATCCCAACCGGTTG 2376
DB GTGTGAGTGTACAGCGCGGATTTGTTGCGGCTTGGCGCAAGACGATCCCAACCGGTTG 1440
QY 2377 ACAGGATCAAACTCTGTGACGATTTGTTGAGTACCGGGGCTGCGGCTGATTTACC 2436
DB ACAGGATCAAACTCTGTGACGATTTGTTGAGTACCGGGGCTGCGGCTGATTTACC 1500
QY 2437 TAGGCGACGAAGTGGGTACGCTCAATGACGACGACTGGTTCGCAAGACGACGAATAGAGCG 2496
DB TAGGCGACGAAGTGGGTACGCTCAATGACGACGACTGGTTCGCAAGACGACGAATAGAGCG 1560

QY 2497 ACACAGCGGTGGGCGCAACCGTCCGGGCTACACAGAGCCCTGTACGCGCAACGCAACG 2556
DB ACACAGCGGTGGGCGCAACCGTCCGGGCTACACAGAGCCCTGTACGCGCAACGCAACG 1620
QY 2557 ATCCGTGACGCGGCGGCAATCTATCAGGCTTGGCGCATATGATTCGCGCTCGCC 2616
DB ATCCGTGACGCGGCGGCAATCTATCAGGCTTGGCGCATATGATTCGCGCTCGCC 1680
QY 2617 AAACCAATCCGCGCTTCCGAGCGCGGCGGCTTGTACATTAATCAACACCAACCAAGCACA 2676
DB AAACCAATCCGCGCTTCCGAGCGCGGCGGCTTGTACATTAATCAACACCAACCAAGCACA 1740
QY 2677 TCATCGGTATCATCCGCAACAAATGCGCTTTTGGCAATTCGCTTCAAGCAATATCCG 2736
DB TCATCGGTATCATCCGCAACAAATGCGCTTTTGGCAATTCGCTTCAAGCAATATCCG 1800
QY 2737 AAACCGTTACCGGCTACACCTGCAAGCGCTTCAAGCGCGACGACCTCATCGGTG 2796
DB AAACCGTTACCGGCTACACCTGCAAGCGCTTCAAGCGCGACGACCTCATCGGTG 1860
QY 2797 GCAAACTGTGCGCTGCAATCAGGATTTGCGCTTCAAGCGCTTCAAGCTCATGCGCTG 2856
DB GCAAACTGTGCGCTGCAATCAGGATTTGCGCTTCAAGCGCTTCAAGCTCATGCGCTG 1920
QY 2857 AATCGCTGACGCGACGCTTCCCAATGCGCTTCAAGCGCTTCAAGCTTTCAGA 2903
DB AATCGCTGACGCGACGCTTCCCAATGCGCTTCAAGCGCTTCAAGCTTTCAGA 1967

RESULT 10
AX300255/c
LOCUS AX300255 795 bp DNA linear PAT 30-NOV-2001
DEFINITION Sequence 81 from Patent WO0185772.
ACCESSION AX300255
VERSION AX300255.1 GI:17381646
KEYWORDS
SOURCE Neisseria meningitidis
ORGANISM Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.

REFERENCE
1
Tang, C.U.
Virulence genes, proteins, and their use
Patent: WO 0185772-A 81 15-NOV-2001;
Microscience Limited (GB)

FEATURES
source
1..795
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"
1..->795
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD12960.1"
/db_xref="GI:17381647"
/translation="MAAFVSCACFMVSCLVVKNVTIGSGRTKIAPLVARDAAVLSA
VLQDIKNLPFDIVFRADFLCAGSISGEVLHRTQTVRDALPKPLLPFRHGGGSGF
PCSDQYVFLDLALISRLPDIIIDILFSGETAVRCVANAOKNGIALLCNHGFHRT
PPRETVCKLQMGEDGADICKAVNMQSADLVLTLLSATLKKAKELAKPIVTVMSQ
TGAIVRAGLVFSSITFGSGTQNSAPQIQVLSALRATLIDCLEAGD"

BASE COUNT 166 a 238 c 223 g 168 t
ORIGIN

Query Match 22.5%; Score 655; DB 6; Length 795;
Best Local Similarity 96.4%; Pred. No. 3.8e-127;
Matches 670; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 GAGTTTGGCTTCCCGAACCGAAGCTGATGCTTGAGCGCAACACCTGTCGCGCAAGCGG 60
DB GAGTTTGGCTTCCCGAACCGAAGCTGATGCTTGAGCGCAACACCTGTCGCGCAAGCGG 669
QY 61 CTGACCGCCCTTTTGTGCCCCCATCGACATCGTAACATCGCTTTGTGGCAAGCTCTTC 120

```
|||||
668 CTGACGCGCCCGCTGCGCCCATCGAATCGTAACATCGTTGGCGCAAGCTCTTTC 609
QY 121 GCTTTGAGCGTGGCAGAAAGCAAAAGTCAGCAGCTCTTCCGCGCTTTCGGCATCACCGCA 180
Db 608 GCTTTGAGCGTGGCAGAAAGCAAAAGTCAGCAGCTCTTCCGCGCTTTCGGCATCACCGCA 549
QY 181 ATTTTCAGATGTCGCGCGCGCGAGTCTCTCATCTGTTTCAGAGCGCATACGATTTCTTCT 240
Db 548 ATTTTCAGATGTCGCGCGCGCGAGTCTCTCATCTGTTTCAGAGCGCATACGATTTCTTCT 489
QY 241 TCGCGCGCGCTGCGGTGAACCTCATGATTCAGAGCAGCGCGCGATGCGGTTTTTTTGA 300
Db 488 TCGCGCGCGCTGCGGTGAACCTCATGATTCAGAGCAGCGCGCGATGCGGTTTTTTTGA 429
QY 301 GCATCGCCACGCGCGCGCGCGCGGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 428 GCATTTGCCCGCGCGCGCGCGCGCGGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
QY 361 TCGCGCAGCGCGCTTCAATCAGCAGTTCGAGCGTTCGAGCGTTCGAGCGTTCGAGCG 420
Db 368 TCGCGCAGCGCGCTTTCGATCAGCGCTCGCGCGTTCGAGCGTTCGAGCGTTCGAGCG 309
QY 421 GCGAACGACGCGCTTTCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 308 GCGAACGACGCGCTTTCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249
QY 481 GCGTGGCAGCGCTGCGCGTGGCGCGATCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 248 GCGTGGCAGCGCTGCGCGTGGCGCGATCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189
QY 541 AAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 188 AAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129
QY 601 GAAAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 128 GAAAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69
QY 661 ATAACCGCTGTTTTTTCGCGTGCAGCGTGCAGCGTGCAGCGTGCAGCGTGCAGCGTGC 695
Db 68 ATAACCGCTATTTTGACACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 34
```

```
RESULT 11
NMA522491/c 329861 bp DNA linear BCT 02-SEP-2002
LOCUS Neisseria meningitidis serogroup A strain Z2491 complete genome,
DEFINITION segment 5/7.
ACCESSION AL162756 AL157959
VERSION AL162756.2 GI:7380091
```

```
Neisseria meningitidis Z2491
Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria, Neisseriales;
Neisseriaceae; Neisseria.
1 (bases 1 to 329861)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,A.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
Nature 404 (6777), 502-506 (2000)
```

```
JOURNAL MEDLINE
PUBMED 10761919
REFERENCE 2 (bases 1 to 329861)
AUTHORS Parkhill,J
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
```

Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).

FEATURES

source
1..329861
/organism="Neisseria meningitidis Z2491"
/mol_type="genomic DNA"
/strain="Z2491"
/db_xref="taxon:122587"
/note="serogroup: A"
194..565
/gene="NMA1439"
194..565
/gene="NMA1439"
/note="NMA1439, possible lipoprotein, len: 123 aa;
unknown, contains a probable N-terminal signal sequence
and an appropriately positioned PS00013 Prokaryotic
membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAB84675.1"
/db_xref="GI:7380092"
/db_xref="SPTREMBL:O9JU92"
/translation="MNKTLSTLPVAILLGGCAAGGNTFGSLDGGTGGGSIIVXMAVE
SOCAELNKSSEWELTALAMSAEKQAEWENKICACVAEPQLTGNQVMDLPDSTR
NQAALATKTVTSACFKLYR"
212..244
/gene="NMA1439"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(571..580)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
592..601
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
638..2047
/gene="thrc"
638..2047
/gene="thrc"
/EC_number="4.2.99.2"
/note="NMA1440, thrc, probable threonine synthase, len:
485 aa; similar to many e.g. SW:THRC.METGL (EMBL:D14071),
thrc, Methylobacillus glyco genes threonine synthase (EC
4.2.99.2) (475 aa), fasta scores; E(): 0, 61.9% identity
in 475 aa overlap. Contains Pfam match to entry PF00291
S.T. dehydratase, Pyridoxal-phosphate dependent enzymes"
/codon_start=1
/transl_table=11
/product="putative threonine synthase"
/protein_id="CAB84676.1"
/db_xref="GI:7380093"
/db_xref="SPTREMBL:O9JU91"
/translation="WKYISRGETAHKPFSEVLMLGLAPDGLMLPEHYPOIGREALD
KWRGLAPELAFEMRLFVTDIPEDDLRDLNRTTEAFGKTEITPRTISDGKIQ
ALSNGPTLAFKDMQAFQNAFVYLNKNGKLNILGATSGDTSAAATLARGKGVN
VFMSPDGMKSAFORAQMSYLDGNHINIAVKGMEDQCQDIVKAVQNDAAFKKHIG
TVNSINGRIVAQVVYTFAGYFKATQSNDEQVFCVPSGNFCAGHIAKQMLPVR
RLIVATNWDLRDFEFTGAYPRNSAHYVTSFSPMDISKASNFPRFVDMRDLPQ
BINTLMAEVAAGKGNLRFALDKVGGKGTSGKSTHADRLATIRQVYEQDELIDPH
TADGVKVARVREERTVLCLETALAAKFDATIREAVGDVAIPRAALEGLENLPRQ
QIVNSADAVNGIIIEQTIA"
875..1780
/gene="thrc"
/note="Pfam match to entry PF00291 S.T. dehydratase,
pyridoxal-phosphate dependent enzymes, score 231.80,
E-value 1e-65"
complement(922..931)
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS

gene	2094..2888 /gene="NMA1441"	2094..2888 /product="hypothetical protein NMA1443"
CDS	2094..2888 /gene="NMA1441" /note="NMA1441, len: 264 aa; unknown"	2094..2888 /protein_id="CAB84679.1" /db_xref="GI:7380096" /translation="MIYKNCNPKVCYAEKMKRMVFIENGCSATVIHRTGIWKIIH IDIIIPPIA" /complement(4509..6980) /gene="NMA1444" complement(4509..6980) /genes="NMA1444" /notes="NMA1444, probable P-type cation-transporting ATPase, len: 823 aa; similar to many e.g. SM:COPA_ENTHR (EMBL:U13292), copA, Enterococcus hirae copA/potassium-transporting ATPase A (EC 3.6.1.36) (727 aa), fasta scores; E(): 0.34, 0% identity in 744 aa overlap. Similar to NMA1539, fasta scores; E(): 0.34, 4% identity in 735 aa overlap. Contains hydrophobic, probable membrane-spanning regions. Contains two Pfam matches to entry PF00122 E1-E2 ATPase, E1-E2 ATPases and PS00154 E1-E2 ATPases phosphorylation site"
stem_loop	2904..2937 /note="stem loop containing DNA uptake sequences: acgat gccgtcgaa gcccttcgagcgc atata"	
misc_feature	2909..2918 /note="Core DNA uptake sequence: gccgtctgaa"	
misc_feature	complement(2923..2932) /label=DUS /note="Core DNA uptake sequence: gccgtctgaa"	
RBS	3065..3068 /label=DUS	
gene	3076..3852 /gene="fpr"	
CDS	3076..3852 /gene="fpr" /EC_number="1.18.1.2" /note="NMA1442, fpr, probable ferredoxin--NADP reductase, len: 258 aa; similar to many e.g. SW:FERK_AZOVI (EMBL:U36319), fpr, Azotobacter vinelandii ferredoxin--NADP reductase (EC 1.18.1.2) (257 aa), fasta scores; E(): 0.69, 8% identity in 255 aa overlap. Similar to NMA1664, fasta scores; E(): 1.3e-30, 34.6% identity in 257 aa overlap. Contains Pfam match to entry PF00175 oxidored_fad, Oxidoreductase FAD/NAD-binding domain"	
misc_feature	codon_start=1 /transl_table=1 /product="putative ferredoxin--NADP reductase"	
misc_feature	protein_id="CAB84678.1" /db_xref="GI:7380095"	
misc_feature	db_xref="SPTREMBL:Q9JRE3" /translation="MAAFNTQKVLVSHVHTDAYFTFCTRDESLRFENGQFVWGLMV DGRPLMAYSVASANWSEHLEFFSIKVDGDLTSRLQHLKVDVLLSKRPTGLVAG DLNPKLXLSTGTGIAPLSITKDPEIYEQFEKILVHGVRYKDLAYDRFTKEL PEHYLGDLVKEKLIYVPIVSRBEFEHRLTDLWVSKLFEIDGLPKINPQDDRAML CGSPALMKTCKVLDLDFGLTVSPKTVGRGVDYLIERAFVQ"	
misc_feature	3394..3768 /gene="fpr"	
misc_feature	note="Pfam match to entry PF00175 oxidored_fad, Oxidoreductase FAD/NAD-binding domain, score 20.00, E-value 8.4e-05"	
misc_feature	3866..3875 /label=DUS /note="Core DNA uptake sequence: gccgtctgaa"	
repeat_unit	3964..3983 /note=">= 90% match to ATTCNNNNNNNGGGAAT"	
repeat_unit	/label=DRS3 complement(3984..4038) /note="RS103"	
repeat_unit	/label=RS103 4039..4058 /note=">= 90% match to ATTCNNNNNNNGGGAAT"	
gene	4159..4320 /label=DRS3	
gene	4159..4320 /gene="NMA1443"	
CDS	4159..4320 /gene="NMA1443" /note="NMA1443, len: 53 aa; unknown, lies within a region of unusually low GC content"	
Query Match	22.5%; Score 655; DB 1; Length 329861;	
Best Local Similarity	96.4%; Pred. No. 4.1e-127;	
Matches	670; Conservative 0; Mismatches 25; Indels 0; Gaps 0;	
Qy	1 GAGTTTGGCTTCCGAAACCGACGTCATGCTTGAGCGGACACCTGTCGGGACGCGG 60	
Db	243562 GAGTTTGGCTTCCGAAACCGACGTCATGCTTGAGCGGACACCTGTCGGGACGCGG 243503	
Qy	61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGTTTGGTGGCAAGCTCTTTC 120	
Db	243502 CTGACCGCCCTTTGCCCCATCGACATCGTAACATCGTTTGGTGGCAAGCTCTTTC 243443	
Qy	121 GCTTGGAGCTGCGAAGAAAGTACGACCTCTTCGCGCTTTGGCGCATACCGCA 180	
Db	243442 GCTTGGAGCTGCGGAAAGAAAGTACGACCTCTTCGCGCTTTGGCGCATACCGCA 243383	
Qy	181 ATTTTCAGATGTCGCGCGGCGAGTCCTCCATCTGTTTCAGAGCGGATAGATTTCTTC 240	
Db	243382 ATTTTCAGATGTCGCGCGGCGAGTCCTCCATCTGTTTCAGAGCGGATAGATTTCTTC 243323	
Qy	241 TCGCGCGGCTGCGGTGAAAATCTCATGTTGCAGAGCGGCGCGATGCGGTTTTTTGA 300	
Db	243322 TCGCGCGGCTGCGGTGAAAATCTCATGTTGCAGAGCGGCGCGATGCGGTTTTTTGA 243263	
Qy	301 GCATGCGGCGCGCGCGCGGCGGCTTTCGCGGAAAAAGCTCGATATCGATATG 360	
Db	243262 GCATTTGCCACGCGCGCGCGGCGGCTTTCGCGGAAAAAGCTCGATATG 243203	
Qy	361 TCGCGGAGGCGGCTTTCGCGCATGCGCGTCTGAAAGTAAACAGAGCGGCTTTCGCGGAC 420	
Db	243202 TCGCGGAGGCGGCTTTCGCGCATGCGCGTCTGAAAGTAAACAGAGCGGCTTTCGCGGAC 243143	
Qy	421 GGAACGAGCGGCTTTCGCGCATGCGCGTCTGAAAGTAAACAGAGCGGCTTTCGCGGAC 480	
Db	243142 GGAACGAGCGGCTTTCGCGCATGCGCGTCTGAAAGTAAACAGAGCGGCTTTCGCGGAC 243083	
Qy	481 GCGTCGCGGAGCGTTCGCGTTCGCGCAATCTTCGCGCATGCTGCGCGCGCATTCGAAA 540	
Db	243082 GCGTCGCGGAGCGTTCGCGTTCGCGCAATCTTCGCGCATGCTGCGCGCGCATTCGAAA 243023	
Qy	541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGCTTTTGTGTTGTTGTTGTTGTTG 600	
Db	243022 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGCTTTTGTGTTGTTGTTGTTGTTG 242963	
Qy	601 GAAAGTACGGGCGATCGCGGGCGACAGCGGCGACCGGCTTTTGTGTTGTTGTTGTTGTTG 660	

Search completed: November 8, 2003, 16:01:00
Job time : 10364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 11:55:01 ; Search time 728 Seconds
(without alignments)
10805.168 Million cell updates/sec

Title: US-09-843-007A-1

Perfect score: 2914

Sequence: 1 GAGTTTCGGTCCCGAAC.....CGTTTCAGCGCATTTCCG 2914

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2914	100.0	2914	21	AAA11732
2	2914	100.0	4173	21	AAA07380
3	2641	90.6	2883	17	AAT09860
4	2641	90.6	2883	17	AAT11179
5	1921.4	65.9	1939	22	AAF61709
6	1921.4	65.9	6878	22	AAF61711
7	1891.4	64.9	1310	22	AAF61710
8	1891.4	64.9	6851	22	AAF61712

C	9	655	22.5	795	24	AAS97236	Neisseria meningit
C	10	639	21.9	765	25	ABX09900	N. meningitidis DN
C	11	639	21.9	48275	21	AAA81501	N. meningitidis pa
C	12	639	21.9	349980	21	AAF21610	Neisseria meningit
C	13	639	21.9	837096	21	AAAG1489	N. meningitidis pa
C	14	637.4	21.9	762	25	ABZ40085	N. gonorrhoeae nuc
C	15	297.2	10.2	29559	23	AAS59546	Propionibacterium
C	16	169	5.8	759	23	AAS52439	E. coli DNA for ce
C	17	161.4	5.5	882	23	AAS56202	Salmonella typhi D
C	18	153.2	5.3	1163020	24	ABQ67197	Listeria innocua C
C	19	153.2	5.3	3011208	24	ABQ69245	Listeria innocua C
C	20	151.6	5.2	799	12	AAQ15228	arod gene. Salmon
C	21	150	5.1	2944528	24	ABA03041	Listeria monocytog
C	22	147.4	5.1	1340	24	ABQ70237	Listeria monocytog
C	23	142.6	4.9	801	23	AAS31122	Enterococcus faeca
C	24	142.6	4.9	6415	20	AAH13212	Enterococcus faeca
C	25	142.6	4.9	6415	24	ABQ99007	Enterococcus faeca
C	26	133	4.6	756	24	ABQ68513	Listeria monocytog
C	27	120.4	4.1	3303	23	AAS54127	Pseudomonas aerugi
C	28	118	4.0	483	24	ABK73108	Bacillus lichenifo
C	29	112.2	3.9	1126	23	AAS29731	Propionibacterium
C	30	111.8	3.8	1794	22	AAH67492	C glutamicum codin
C	31	111.8	3.8	1981	24	ABS85343	DNA encoding C. gl
C	32	111.8	3.8	349980	22	AAH68531	C glutamicum codin
C	33	111.8	3.8	349980	22	AAH68532	C glutamicum codin
C	34	104.6	3.6	1102	22	AAF71764	Corynebacterium gl
C	35	104.4	3.6	1704	17	AAT11413	Sequence encoding
C	36	99.4	3.4	4403765	22	AAI99683	Mycobacterium tube
C	37	99.4	3.4	4411529	22	AAI99682	Mycobacterium tube
C	38	91.8	3.2	1113	23	AAS9702	Propionibacterium
C	39	85.8	2.9	22934	23	AAS59613	Propionibacterium
C	40	82.2	2.8	293	24	ABN21665	Human ORFX polynuc
C	41	76.2	2.6	349980	24	ABQ81845	Bifidobacterium lo
C	42	73.6	2.5	3600	17	AAT13227	Thermostable enzym
C	43	65	2.2	1689	24	ABK73366	Bacillus lichenifo
C	44	61.6	2.1	2238	23	AAS2760	DNA encoding novel
C	45	59.8	2.1	1782	16	AAT03251	Pseudomonas mesoac

ALIGNMENTS

RESULT 1

AAA11732
ID AAA11732 standard; DNA; 2914 BP.

XX AAA11732;

XX 21-JUL-2000 (first entry)

DE N. polysaccharea amylsucrase DNA.

XX Alpha-1,6-branched alpha-1,4-glucan; branching enzyme; binder; carrier;
KW transgenic plant; flavor; perfume; packaging material; papermaking;
KW ultra-violet light adsorber; starch; textile; wetting agent;
KW amylsucrase; ds.

XX Neisseria polysaccharea.

XX Key Location/Qualifiers

FT CDS 957..2867

FT /*tag=

FT /product= "amylsucrase"

XX WO2000022140-A1.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-EP07562.

XX 09-OCT-1998; 98DE-1046635.

XX 27-MAY-1999; 99DE-1024342.

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX Buettcher V, Quanz M;
 PT WPI; 2000-317992/27.
 DR P-PSDB; AAW90979.

XX New nucleic acid encoding a branching enzyme, useful for in vitro
 PT synthesis of branched glucans and to prepare transgenic plants
 PT producing modified starch -
 XX Disclosure; Page 99-102; 115pp; German.

XX This invention describes a novel nucleic acid (I) isolated from
 CC Neisseria which encodes a branching enzyme (II). (I) is used for
 CC recombinant production of (II) subsequently used in the in vitro
 CC production of alpha-1,6-branched alpha-1,4-glucans. It is also used to
 CC prepare transgenic plants that produce starches with modified properties.
 CC (III) are used as binders for tablets, carriers for pharmaceuticals,
 CC flavors and perfumes and powdered additives, packaging materials,
 CC ultra-violet light adsorbers in sunscreens and also for any of the usual
 CC applications of starch in foods, papermaking, as textile size, in soil
 CC stabilization, as wetting agent for agricultural chemicals, as polymer
 CC additives etc. Fragments of (I) are useful as PCR primers and antisense
 CC molecules or ribozymes for inhibiting expression of (I), and the
 CC regulatory region of (II) can be used to control expression of
 CC heterologous sequences in host cells. (I) provides an inexpensive method
 CC for producing alpha-1,6-branched alpha-1,4-glucans (III), producing
 CC products that can be tailored for particular applications, particularly
 CC by controlling the degree of branching. Starch from transgenic plants
 CC has increased gel strength; reduced phosphate content; reduced peak
 CC viscosity; lower pasting temperature and granule size and/or altered
 CC sidechain distribution. This sequence encodes an amylase isolated
 CC from Neisseria polysacchara which is described in the method of the
 CC invention.

XX Sequence 2914 BP; 662 A; 868 C; 773 G; 611 T; 0 other;

Query Match 100.0%; Score 2914; DB 21; Length 2914;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2914; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTTTGGTTCCTCCGAAACGGAACGATGATGCTTGGAGCGGAAACCTGTCCGCGCAAGGGCG 60
 DB 1 GAGTTTGGTTCCTCCGAAACGGAACGATGATGCTTGGAGCGGAAACCTGTCCGCGCAAGGGCG 60
 QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACAAATCGTTTGGTGGCAAGCTCTTTC 120
 DB 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACAAATCGTTTGGTGGCAAGCTCTTTC 120
 QY 121 GCTTTGAGCGTGGCAGAAAGCAAGTCAAGCTCTTTCGCGTTTGGCGCATCACCGCA 180
 DB 121 GCTTTGAGCGTGGCAGAAAGCAAGTCAAGCTCTTTCGCGTTTGGCGCATCACCGCA 180
 QY 181 ATTTTCAGATGTCGCGCGGCGAGTCTCCATCTGTTTCAGACGGCATACGATTTCTTC 240
 DB 181 ATTTTCAGATGTCGCGCGGCGAGTCTCCATCTGTTTCAGACGGCATACGATTTCTTC 240
 QY 241 TCGCGGGCGTGGGTGAACCTCATGATTCGACAGCGCGCGGATCGGTTTTCGA 300
 DB 241 TCGCGGGCGTGGGTGAACCTCATGATTCGACAGCGCGCGGATCGGTTTTCGA 300
 QY 301 GCATGCGCCACGCGCGCGGAGCGGCGGTTTTCGCGGAAAGCTTCGATATCGATAATG 360
 DB 301 GCATGCGCCACGCGCGCGGAGCGGCGGTTTTCGCGGAAAGCTTCGATATCGATAATG 360
 QY 361 TCGCGGGCGGCTTTCATTCAGCGAGTTCGAGTTCGAATATTAATTCGCGAAGC 420
 DB 361 TCGCGGGCGGCTTTCATTCAGCGAGTTCGAGTTCGAATATTAATTCGCGAAGC 420
 QY 421 GCGAAGCAGCGCGCTTCGCGCGTCTGAAAGCTAAACAGCAGCGCGCTTGTCCGCGAGC 480

DB 421 GCGAAGCAGCGCGCTTCGCGCATGCGGTCTGAAACGTAACAGCAGCGCGCTTGTCCGCGAGC 480
 QY 481 GCGTCGCGGAGCGGTCTCGGTGTGCGCAATCTTTCGCGCATGCTGCGCGCATTTCCAAA 540
 DB 481 GCGTCGCGGAGCGGTCTCGGTGTGCGCAATCTTTCGCGCATGCTGCGCGCATTTCCAAA 540
 QY 541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTTGTATTTGGTCAAGTACGGCG 600
 DB 541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTTGTATTTGGTCAAGTACGGCG 600
 QY 601 GAATGACGCGGCGCATCGCGCGGCGACAGCGGCGATTTTGGTGGCGTCCGCTTCG 660
 DB 601 GAATGACGCGGCGCATCGCGCGGCGACAGCGGCGATTTTGGTGGCGTCCGCTTCG 660
 QY 661 ATACGCGTGTTTTTCAGCGCTCGTGTGCTATGCGGTGTGCGGTGTGCGGTGTGAAAGGAAC 720
 DB 661 ATACGCGTGTTTTTCAGCGCTCGTGTGCTATGCGGTGTGCGGTGTGCGGTGTGAAAGGAAC 720
 QY 721 GGTAAAGCGCAATATAGCAAGGCGACAGCGATGTTTCAGACGGCATTTCTGTGGCGG 780
 DB 721 GGTAAAGCGCAATATAGCAAGGCGACAGCGATGTTTCAGACGGCATTTCTGTGGCGG 780
 QY 781 CGGCTTGATATGAATCAAGCAGCATCGCATATCGGAATCGAGACTTTGGCAAGCCCTG 840
 DB 781 CGGCTTGATATGAATCAAGCAGCATCGCATATCGGAATCGAGACTTTGGCAAGCCCTG 840
 QY 841 TCTTTCTAGTCAGTCGCGAGTCTTTCAGTATGATTCAGTCAAGCAAGCCCTACAGCGCAT 900
 DB 841 TCTTTCTAGTCAGTCGCGAGTCTTTCAGTATGATTCAGTCAAGCAAGCCCTACAGCGCAT 900
 QY 901 TTGAGGATAGCGCGGCGAGACCGCGGTGCGAAACTTTCAGAAATCGAGCAGCGCATCATGT 960
 DB 901 TTGAGGATAGCGCGGCGAGACCGCGGTGCGAAACTTTCAGAAATCGAGCAGCGCATCATGT 960
 QY 961 TGACCCCGCAGCAGCAAGTCTGTTGATTTTACAGTACCTCAAAACAGCGCATTTGGACA 1020
 DB 961 TGACCCCGCAGCAGCAAGTCTGTTGATTTTACAGTACCTCAAAACAGCGCATTTGGACA 1020
 QY 1021 TCTACAGCGCGGAGCAGCGCGCGCATCGGAAATCGAAGACTTGGCGCGAGTTTTCGC 1080
 DB 1021 TCTACAGCGCGGAGCAGCGCGCGCATCGGAAATCGAAGACTTGGCGCGAGTTTTCGC 1080
 QY 1081 GCGCATGATGATGCGATTTTCCCAAACTGATGATGATGATGATGATGATGATGATGATGAT 1140
 DB 1081 GCGCATGATGATGCGATTTTCCCAAACTGATGATGATGATGATGATGATGATGATGATGAT 1140
 QY 1141 ACGAGCGCTGCTGCTATGCTGGAATGCTGCTGCGCGAGCATGCGAAGCTATTCCC 1200
 DB 1141 ACGAGCGCTGCTGCTATGCTGGAATGCTGCTGCGCGAGCATGCGAAGCTATTCCC 1200
 QY 1201 AACGCAACTCATCTCTTAAAGATATCGATATCGCGCGCGGAAACAAACCCGAGTGGATT 1260
 DB 1201 AACGCAACTCATCTCTTAAAGATATCGATATCGCGCGCGGAAACAAACCCGAGTGGATT 1260
 QY 1261 TGTCCAAACAAAGTTCGCGCGGCGTGTGCTAGCTGATTTGCTTTGCGCGGATTTGAAGG 1320
 DB 1261 TGTCCAAACAAAGTTCGCGCGGCGTGTGCTAGCTGATTTGCTTTGCGCGGATTTGAAGG 1320
 QY 1321 GCTTGAAGATATAAATTCCTTTTAAAGAGCTTGTGTTTTCATTTATCTGACCTGTATGC 1380
 DB 1321 GCTTGAAGATATAAATTCCTTTTAAAGAGCTTGTGTTTTCATTTATCTGACCTGTATGC 1380
 QY 1381 GCTGTTTAAATGCTCGAAGCAAAAGCAGCGCGGCTATGCGGTGAGCAGCTACCGG 1440
 DB 1381 GCTGTTTAAATGCTCGAAGCAAAAGCAGCGCGGCTATGCGGTGAGCAGCTACCGG 1440
 QY 1441 ATGTCAATTCGCGCATCTGGGCAAAATAGCGGCTTTCGCGGAGTCAFTTGTGCGCTGACG 1500
 DB 1441 ATGTCAATTCGCGCATCTGGGCAAAATAGCGGCTTTCGCGGAGTCAFTTGTGCGCTGACG 1500
 QY 1501 AAGCGCGCATTTCCGCGCGTGTGCTGATTTTATCTTCAAACACCTTCCAAACAGCAAGT 1560
 DB 1501 AAGCGCGCATTTCCGCGCGTGTGCTGATTTTATCTTCAAACACCTTCCAAACAGCAAGT 1560

QY 1 GAGTTTGGTTCCTCCGAAACGACGATGATGCTTGAGCCGAAACACCTGTCTCCGCAAGGCG 60
DB |||||
QY 1015 GAGTTTGGTTCCTCCGAAACGACGATGATGCTTGAGCCGAAACACCTGTCTCCGCAAGGCG 1074
DB |||||
QY 61 CTGACCGCCCTTTTGGCCCAATGCAATCGTAACAAATCGGTTTGGTGCACAGCTTTTC 120
DB |||||
QY 1075 CTGACCGCCCTTTTGGCCCAATGCAATCGTAACAAATCGGTTTGGTGCACAGCTTTTC 1134
DB |||||
QY 121 GCTTTGAGGCTGCGAGAAAGCAAGTCAGCACTGCTTCCTCCGCTTTCGCGCTTTCGCGCATCACCGCA 180
DB |||||
QY 1135 GCTTTGAGGCTGCGAGAAAGCAAGTCAGCACTGCTTCCTCCGCTTTCGCGCTTTCGCGCATCACCGCA 1194
DB |||||
QY 181 AATTTGAGATGTCGCGCGCGAGTCTCTCCATCTGTTTCAGAGCGGATACGATTTCTCT 240
DB |||||
QY 1195 AATTTGAGATGTCGCGCGCGAGTCTCTCCATCTGTTTCAGAGCGGATACGATTTCTCT 1254
DB |||||
QY 241 TGGCGGGGCTGGGTGAACCTCATGATTCAGAGCAGGCGCGGATGCGGTTTTCG 300
DB |||||
QY 1255 TGGCGGGGCTGGGTGAACCTCATGATTCAGAGCAGGCGCGGATGCGGTTTTCG 1314
DB |||||
QY 301 GCATGCGCACGCGCGCGAGCGGCTTTCGCGGAAAGCAAGCTTCGATATCGATAATG 360
DB |||||
QY 1315 GCATGCGCACGCGCGCGAGCGGCTTTCGCGGAAAGCAAGCTTCGATATCGATAATG 1374
DB |||||
QY 361 TGGCGAGCGGCTTTCATTCAGAGCTGAGCAGTTCAAAATAAATTCGTCGGAACAC 420
DB |||||
QY 1375 TGGCGAGCGGCTTTCATTCAGAGCTGAGCAGTTCAAAATAAATTCGTCGGAACAC 1434
DB |||||
QY 421 GGAACGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 480
DB |||||
QY 1435 GGAACGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 1494
DB |||||
QY 481 GGTTCGCGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 540
DB |||||
QY 1495 GGTTCGCGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 1554
DB |||||
QY 541 AATTCGCGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 600
DB |||||
QY 1555 AATTCGCGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 1614
DB |||||
QY 601 GAAAGTACGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 660
DB |||||
QY 1615 GAAAGTACGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 1674
DB |||||
QY 661 AATTCGCGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 720
DB |||||
QY 1675 AATTCGCGAGCGGCTTTCGCGCATGCTGCAAGTAAACAGCAGCGCTTTCGCGAGC 1734
DB |||||
QY 721 GGTAAAGACGCAATTATAGCAAGGCAAGCGCAATGTTTCAGAGCGGCTTTCGCGAGC 780
DB |||||
QY 1735 GGTAAAGACGCAATTATAGCAAGGCAAGCGCAATGTTTCAGAGCGGCTTTCGCGAGC 1794
DB |||||
QY 781 CGGCTTGAATGAAATCAAGCAGCATTCGCGATATCGGAATGCAAGCTTTCGCGAGC 840
DB |||||
QY 1795 CGGCTTGAATGAAATCAAGCAGCATTCGCGATATCGGAATGCAAGCTTTCGCGAGC 1854
DB |||||
QY 841 TCTTTTCTAGTCAGTCGCGAGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 900
DB |||||
QY 1855 TCTTTTCTAGTCAGTCGCGAGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 1914
DB |||||
QY 901 TCGAGGATACGCGCGAGCAGCGGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 960
DB |||||
QY 1915 TCGAGGATACGCGCGAGCAGCGGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 1974
DB |||||
QY 961 TGACCCCGCAGCAGGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 1020
DB |||||
QY 1975 TGACCCCGCAGCAGGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 2034
DB |||||
QY 1021 TCTACGCGCGAGCAGCGGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 1080
DB |||||
QY 2035 TCTACGCGCGAGCAGCGGCTTTCGCGATGATGATTCAGAGCGGCTTTCGCGAGC 2094
DB |||||
QY 1081 GCCGATGGATACGCAATTTCCCAAACTGATGAGAGGCTTCGAGCGCTGTCAGGCAACA 1140
DB |||||

DB 2095 GCCGATGGATACGCAATTTCCCAAACTGATGAGAGGCTTCGAGCGCTGTCAGGCAACA 2154
QY |||||
DB 1141 ACGAAGCCCTCTGCTGCTATGCTGGAATGCTGCTGCGAGGCTGCGCAAGCTTATTC 1200
QY |||||
DB 2155 ACGAAGCCCTCTGCTGCTATGCTGGAATGCTGCTGCGAGGCTGCGCAAGCTTATTC 2214
QY |||||
DB 1201 AACGCAACTCATCTTTAAAGATATCGATATCGCGCGCGGCAAAACCAACCCCGATTTG 1260
DB |||||
DB 2215 AACGCAACTCATCTTTAAAGATATCGATATCGCGCGCGGCAAAACCAACCCCGATTTG 2274
QY |||||
DB 1261 TGTCCAAACAAACAGTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY |||||
DB 2275 TGTCCAAACAAACAGTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2334
QY |||||
DB 1321 GCTTCGAAGATATAAATTCCTTATTTCAAGAGCTTGGTTTGACTTATTCGACCTGATGC 1380
DB |||||
DB 2335 GCTTCGAAGATATAAATTCCTTATTTCAAGAGCTTGGTTTGACTTATTCGACCTGATGC 2394
QY |||||
DB 1381 CGCTGTTTAAATGCTCCTGAAAGCAAGAGCGGCTTATGCGGTGAGCAGCTACCGG 1440
DB |||||
DB 2395 CGCTGTTTAAATGCTCCTGAAAGCAAGAGCGGCTTATGCGGTGAGCAGCTACCGG 2454
QY |||||
DB 1441 ATGTCAATTCGCGCATGTCGCGCAATAGAGCACTTTCGCGGAGTCAATTCGCTGCTGCA 1500
DB |||||
DB 2455 ATGTCAATTCGCGCATGTCGCGCAATAGAGCACTTTCGCGGAGTCAATTCGCTGCTGCA 2514
QY |||||
DB 1501 AAGCCGCGATTTCCGCGCTGCTGCTTATCTTCAACACACCTTCCAAACGAGCAAT 1560
DB |||||
DB 2515 AAGCCGCGATTTCCGCGCTGCTGCTTATCTTCAACACACCTTCCAAACGAGCAAT 2574
QY |||||
DB 1561 GGGCGCAAGCTGCGCGCGGCGGCGGCTTTCGAGCAATTTCTACTATATTTTCCCG 1620
DB |||||
DB 2575 GGGCGCAAGCTGCGCGCGGCGGCGGCTTTCGAGCAATTTCTACTATATTTTCCCG 2634
QY |||||
DB 1621 ACCGCGGATGTCGCGCAATACGACCGCACCTTCGCGGAAATCTTCCCGACGACAC 1680
DB |||||
DB 2635 ACCGCGGATGTCGCGCAATACGACCGCACCTTCGCGGAAATCTTCCCGACGACAC 2694
QY |||||
DB 1681 CGGCGGCTTCTCGCAATGGAAGAGGAGCGCTGCTGAGGAGCTTCAATTCCTTCC 1740
DB |||||
DB 2695 CGGCGGCTTCTCGCAATGGAAGAGGAGCGCTGCTGAGGAGCTTCAATTCCTTCC 2754
QY |||||
DB 1741 AATGGACCTTGAATTAACAGCAACCGCTGCTTATTCGCGCAATGCGCGGCAATTCGCT 1800
DB |||||
DB 2755 AATGGACCTTGAATTAACAGCAACCGCTGCTTATTCGCGCAATGCGCGGCAATTCGCT 2814
QY |||||
DB 1801 TCTTTCGCAACTTGGCGGCTTGAATCTGCTGCTGCTGCTGCTGCTTATTTGGAAC 1860
DB |||||
DB 2815 TCTTTCGCAACTTGGCGGCTTGAATCTGCTGCTGCTGCTGCTTATTTGGAAC 2874
QY |||||
DB 1861 AAATGGGCAACAGCTGCGGAAACCTGCGGAGCGGCAAGCTTCAATTCGCGGCTTCAATG 1920
DB |||||
DB 2875 AAATGGGCAACAGCTGCGGAAACCTGCGGAGCGGCAAGCTTCAATTCGCGGCTTCAATG 2934
QY |||||
DB 1921 CGGTTATGCTATGCTGCGCGGCTGCTTCTTCAATTCGAGGCTTTCGCGGCTTCAATG 1980
DB |||||
DB 2935 CGGTTATGCTATGCTGCGCGGCTGCTTCTTCAATTCGAGGCTTTCGCGGCTTCAATG 2994
QY |||||
DB 1981 ACCAGCTTCTCAATTAACATCGGCGAGGAGGAGTTCGCTTCAACCTTCAACCTTCAAC 2040
DB |||||
DB 2995 ACCAGCTTCTCAATTAACATCGGCGAGGAGGAGTTCGCTTCAACCTTCAACCTTCAAC 3054
QY |||||
DB 2041 TGGCATTTGTTGGAACACCTTTCGCGGAGTTCGCTTCAACCTTCAACCTTCAACCTTCAAC 2100
DB |||||
DB 3055 TGGCATTTGTTGGAACACCTTTCGCGGAGTTCGCTTCAACCTTCAACCTTCAACCTTCAAC 3114
QY |||||
DB 2101 CTTACCGGCAACCTGCGCGGAGTTCGCTTCAACCTTCAACCTTCAACCTTCAACCTTCAAC 2160
DB |||||
DB 3115 CTTACCGGCAACCTGCGCGGAGTTCGCTTCAACCTTCAACCTTCAACCTTCAACCTTCAAC 3174
QY |||||
DB 2161 ACATCGGCTGAGGCTTTCGCGGAGTTCGCTTCAACCTTCAACCTTCAACCTTCAACCTTCAAC 2220
DB |||||

```
Db 3175 ACATCGGCTGAGCGTTTCCCGATGAGAGCGCGGCATATCTGGGCATAGAGCGGCTACGACC 3234
Qy 2221 ACCGCCAATTCCTCAACCGCTTCTTCGTCAACCGCTTTCGACGGCAGCTTCGTCTGTGGCG 2280
Db 3235 ACCGCCAATTCCTCAACCGCTTCTTCGTCAACCGCTTTCGACGGCAGCTTCGTCTGTGGCG 3294
Qy 2281 TACCGCTTCGATACCAACGACGAGCGAGCTCCGCTGTCAGTGTGTACAGCGCGGCAT 2340
Db 3295 TACCGCTTCGATACCAACGACGAGCGAGCTCCGCTGTCAGTGTGTACAGCGCGGCAT 3354
Qy 2341 TGGTCGGCTTGGCGCAAGACGATCCGCCAGCGCTTGGACGCGATCAAACTCTTGTACAGCA 2400
Db 3355 TGGTCGGCTTGGCGCAAGACGATCCGCCAGCGCTTGGACGCGATCAAACTCTTGTACAGCA 3414
Qy 2401 TTGCTTTGATACCGCGCGCTTCGCGCTGATTTACTTAGCGGACGAGTGGTAGCGTCA 2460
Db 3415 TTGCTTTGATACCGCGCGCTTCGCGCTGATTTACTTAGCGGACGAGTGGTAGCGTCA 3474
Qy 2461 ATGACGACGACTGTGCGCAAGACGACATTAAGACGACGACGACGACGACGACGACGACG 2520
Db 3475 ATGACGACGACTGTGCGCAAGACGACATTAAGACGACGACGACGACGACGACGACGACG 3534
Qy 2521 CGCGTACCAAGAACCCCTGTACGCGCAACGCAACGATCCGTCGACCGAGCCGGGAAA 2580
Db 3535 CGCGTACCAAGAACCCCTGTACGCGCAACGCAACGATCCGTCGACCGAGCCGGGAAA 3594
Qy 2581 TCTATCAGGCTTGGCGCATATGATTGCGCTCCGCCAAAGCAATCCGCGCTTCGACGCG 2640
Db 3595 TCTATCAGGCTTGGCGCATATGATTGCGCTCCGCCAAAGCAATCCGCGCTTCGACGCG 3654
Qy 2641 GCAGCTGGTTACATTCACACCAACCAACGACATCATCGGCTACATTCGCGCAACATG 2700
Db 3655 GCAGCTGGTTACATTCACACCAACCAACGACATCATCGGCTACATTCGCGCAACATG 3714
Qy 2701 CGCTTTGGCATTCGTAACCTTACGGATATTCGCAACCGTTCACCGGATACCCCTGC 2760
Db 3715 CGCTTTGGCATTCGTAACCTTACGGATATTCGCAACCGTTCACCGGATACCCCTGC 3774
Qy 2761 AAGCCATGCCCTTCAAGGCGACGACCTCATCGCTGGCAAACTGTGACGCTGATCAGG 2820
Db 3775 AAGCCATGCCCTTCAAGGCGACGACCTCATCGCTGGCAAACTGTGACGCTGATCAGG 3834
Qy 2821 ATTGACGCTTACGCGCTTACGCTATGCTGATGCTGCTGAAATCGCCTGACGACGCTTCCA 2880
Db 3835 ATTGACGCTTACGCGCTTACGCTATGCTGATGCTGCTGAAATCGCCTGACGACGCTTCCA 3894
Qy 2881 AATGCGCTGTAACCGTTTTCAGACGCGATTTCG 2914
Db 3895 AATGCGCTGTAACCGTTTTCAGACGCGATTTCG 3928
```

RESULT 3

AAT09860

ID AAT09860 standard; DNA; 2883 BP.

XX AC AAT09860;

XX DT 08-APR-1996 (first entry)

XX DE Neisseria polysaccharea amylase DNA sequence.

XX KW Amylosucrase; enzyme; amylose, linear 1,4-glucan; transgenic plant;

XX KW EC-2.4.1.4; crop improvement; ss.

XX OS Neisseria polysaccharea.

XX FH Key Location/Qualifiers

XX FT 5'UTR 1..938

XX FT sig_peptide 939..986

XX FT CDS 939..2780

XX FT CDS /*tag= c

```
XX W09531553-A1.
XX 23-NOV-1995.
XX 18-MAY-1995; 95WO-EP01893.
XX 22-DEC-1994; 94DB-4447388.
XX 18-MAY-1994; 94DE-4417879.
XX (GENB-) INST GENBIOLOGISCHE FORSCHUNG.
XX Buettcher V, Kossmann J, Welsh T;
XX WPI; 1996-010938/01.
XX P-FSD; AAR8386.
XX DNA encoding amylo:sucrose from Neisseria polysaccharea - for prodn.
XX of linear 1,4-glucan(s), esp. amylose, from sucrose.
XX Claim 1; Page 39-42; 56pp; English.
XX This DNA sequence encodes an amylase which allows the
XX synthesis of linear alpha-1,4-glucans from the substrate sucrose
XX by bacteria, fungi and plants, or in cell-free systems. This
XX sequence may be expressed recombinantly.
XX Sequence 2883 BP; 661 A; 852 C; 758 G; 612 T; 0 other;
```

Query Match 90.6%; Score 2641; DB 17; Length 2883;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 2852; Conservative 0; Mismatches 30; Indels 33; Gaps 16;

Qy 1 GAGTTTTCGCTTCCCGAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 60

Db 1 GAGTTTTCGCTTCCCGAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 60

Qy 61 CTGACCGGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 120

Db 61 -----ACCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 114

Qy 121 GCTTTGACGCTGCGCAAGCAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 180

Db 115 GCTTTGACGCTGCGCAAGCAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 173

Qy 181 ATTTGACGATGCTCGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 240

Db 174 ATTTGACGATGCTCGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 233

Qy 241 TCGCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 300

Db 234 TCGCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 290

Qy 301 GCATCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 360

Db 291 GCATCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 344

Qy 361 TCGCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 420

Db 345 TCGCGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 404

Qy 421 GCGCAACGAGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 480

Db 405 GCGCAACGAGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 464

Qy 481 CGCTCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 540

Db 465 CGCTCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 524

Qy 541 AAATCGCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 600

Db 525 AAATCGCGCGCGCTGCGCGCTTTCGCGCAACCGAAGCTGATGCTTTCAGCGCAACACCTGTCCGCAAGCGG 584

QY 601 GAAAGTACCGCGGCATCGCGGCGACACAGCGGCA CGGCGAATTTGGTGGTCCGCTTCGC 660
DB 585 GAAAGTACCGCGGCATCGCGGCGACACAGCGGCA CGGCGAATTTGGTGGTCCGCTTCGC 644
QY 661 ATAAAGCGTCTTTTGAAGCGTCA -GGCTGGTGGCATGGCGGTGTTGGCGCTGAAAGGAA 719
DB 645 ATAAAGCGTCTTTTGAAGCGTCAAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 704
QY 720 CGGTAAAGACGCAATTTATAGCAAGGCA CAGGCAATTTTTCAGACGGCAATTTCTGTGGG 779
DB 705 CGGTAAAGACGCAATTTATAGCAAGGCA CAGGCAATTTTTCAGACGGCAATTTCTGTGGG 764
QY 780 CGGCTTGATATGAATCAAGCAGCATCCGATATCGGAATTCAGACTGGGCAAGCGCT 839
DB 765 CGGCTTGATATGAATCAAGCAGCATCCGATATCGGAATTCAGACTGGGCAAG -CCT 823
QY 840 GTCTTTTCAGTCAAGTCCGAGTTCCTGAGTATGATTTGACGACAGCGCTTACAGGCA 899
DB 824 GTCTTTTCAGTCAAGTCCGAGTTCCTGAGTATGATTTGACGACAGCGCTTACAGGCA 883
QY 900 TTTGCMAGATACGCGGCGAGACCGCGGTGCGAACTTCAGAACTGGAGCAGGCAATCATG 959
DB 884 TTTGCMAGATACGCGGCGAGACCGCGGTGCGAACTTCAGAACTGGAGCAGGCAATCATG 941
QY 960 TTTGACCCCGACGAGCAAGTCCGATTTGATTTTACAGTACCTCAAAACAGCATCTTGGAC 1019
DB 942 TTTGACCCCGACGAGCAAGTCCGATTTGATTTTACAGTACCTCAAAACAGCATCTTGGAC 1001
QY 1020 ATCTACACCGCGGAA CAGCGCGCGGATCGGAAATTCGAAAGCTGGCGGAGTTTCG 1079
DB 1002 ATCTACACCGCGGAA CAGCGCGCGGATCGGAAATTCGAAAGCTGGCGGAGTTTCG 1061
QY 1080 CGCGGCAATGATACGATTTCCGCAACTGATGAAAGCTGACAGAGCTGACGCGCAAC 1139
DB 1062 CGCGGCAATGATACGATTTCCGCAACTGATGAAAGCTGACAGAGCTGACGCGCAAC 1121
QY 1140 AACGAAGCCCTGCTGCTATGCTGGAATGCTGCTGGCGAGGCAATGGCAAGCTATTCC 1199
DB 1122 AACGAAGCCCTGCTGCTATGCTGGAATGCTGCTGGCGAGGCAATGGCAAGCTATTCC 1181
QY 1200 CAAAGCAATCACTCTTAAAGATATCGATATCGCGGCGGAAACAAACCCCGATGGATT 1259
DB 1182 CAAAGCAATCACTCTTAAAGATATCGATATCGCGGCGGAAACAAACCCCGATGGATT 1241
QY 1260 TTGTCCAAACAAACAGTCCGCGCGTGTCTAGTTGATTTGTTTCCGCGCAATTTGAAG 1319
DB 1242 TTGTCCAAACAAACAGTCCGCGCGTGTCTAGTTGATTTGTTTCCGCGCAATTTGAAG 1301
QY 1320 GCTTGAAAGATAAATTCCTTATTTTCAAGAGCTTGGTTGACTTATCTGACAGCTGATG 1379
DB 1302 GCTTGAAAGATAAATTCCTTATTTTCAAGAGCTTGGTTGACTTATCTGACAGCTGATG 1361
QY 1380 CCGCTGTTTAAATGCCCTGAAGCAAAAGCGAGCGGCTATGCGGTACAGCAGCTACCGC 1439
DB 1362 CCGCTGTTTAAATGCCCTGAAGCAAAAGCGAGCGGCTATGCGGTACAGCAGCTACCGC 1421
QY 1440 GATGTCATTCGGCACTGGGCA CAATAGGAGCTTGGCGGAGTCAATTCGCTGGCTGCAAC 1499
DB 1422 GATGTCATTCGGCACTGGGCA CAATAGGAGCTTGGCGGAGTCAATTCGCTGGCTGCAAC 1481
QY 1500 GAAGCGGCAATTTCCGCGTGTGATTTTATCTTCAACCAACCTTCCCAACGAAACAGAA 1559
DB 1482 GAATC - GATTTTCCGCGTGTGATTTTATCTTCAACCAACCTTCCCAACGAAACAGAA 1539
QY 1560 TGGCGCAAGCTGCGCGCGCGGAGCGCGCTTTTCGAAATTTCTATATTTTTCCTCC 1619
DB 1540 T -GGCGCAAGCTG - - -CGCGGCGAGCGCGCTTTTCGAAATTTCTATATTTTTCCTCC 1595
QY 1620 GACGCGGATTCGCGGCAATAGCAGCGCGCTGCGGAGTCTTCCCGGAGCAGGAC 1679
DB 1596 GACGCGGATTCGCGGCAATAGCAGCGCGCTGCGGAGTCTTCCCGGAGCAGGAC 1655
QY 1680 CCGCGGCGCTTCTCGCAATGGAAGACGAGCGCTGGGTGAGCAGCTTCAATTCCTTC 1739

DB 1656 CCGCGCGCTTCTCGCAACTGGAAGACGAGCGCTGGGTGTGGACGACCTTCAATTCCTTC 1715
QY 1740 CAATGGGACCTTTGAATTTACAGCAACCCCGTGGTATTTCCGCGCAATTTGGGCGGCAATGCTG 1799
DB 1716 CAATGGGACCTTTGAATTTACAGCAACCCCGTGGTATTT -CGCGCAATTTGGGCGGCAATGCTG 1774
QY 1800 TTCTTGCCAACTTTGGGCTTTGACATCTGAGATGATGATGGGTTCCTTTATTTTGGAA 1859
DB 1775 TTCTTGCCAACTTTGGGCTTTGACATCTGAGATGATGATGGGTTCCTTTATTTTGGAA 1834
QY 1860 CAATGGGACAGCTGCAAAAACCTTCCGCGAGCGGCAACGCCCTCATCCGCGGCTTCAAT 1919
DB 1835 CAATGGGACAGCTGCAAAAACCTTGC - -CGAGCGCAACGCCCTCATCCGCGGCTTCAAT 1892
QY 1920 GCGGTATGCGTATTTGGCGCGCGCGCGGTGTTCTTCAATTCGGAAGCGCATCTCCACCC 1979
DB 1893 GCGGTATGCGTATTTGGCGCGCGCGCGGTGTTCTTCAATTCGGAAGCGCATCTCCACCC 1952
QY 1980 GACCAAGTCTCCAAATACATCGGCGAGGACGAAATGCAAAATCGGTTTCAAAACCCCTGCA 2039
DB 1953 GACCAAGTCTCCAAATACATCGGCGAGGACGAAATGCAAAATCGGTTTCAAAACCCCTGCA 2012
QY 2040 ATGCAATTTGTTGGAACACCCCTTTGCCACGCGGAGGTTCAACCTGCTTCAATGAGCGCTG 2099
DB 2013 ATGCAATTTGTTGGAACACCCCTTTGCCACGCGGAGGTTCAACCTGCTTCAATGAGCGCTG 2072
QY 2100 ACTTACCGCAACCTTGCAGGATACATCGGCTGGTCAACTAGTCCGCGAGCAGCAGC 2159
DB 2073 ACTTACCGCAACCTTGCAGGATACATCGGCTGGTCAACTAGTCCGCGAGCAGCAGC 2132
QY 2150 GACATCGGCTGAGCTTTGCGGATGAAAGACCGCGGATATCTGGGATTAAGCGGCTTACAC 2219
DB 2133 GACATCGGCTGAGCTTTGCGGATGAAAGACCGCGGATATCTGGGATTAAGCGGCTTACAC 2192
QY 2220 CACGCGCAATTTCTTCAAGCGCTTCTTGTGCAACCGTTTTCAGCGGAGCTTCCGCTGTCG 2279
DB 2193 CACGCGCAATTTCTTCAAGCGCTTCTTGTGCAACCGTTTTCAGCGGAGCTTCCGCTGTCG 2252
QY 2280 GTACGCTTTCAAATCAACCCCAAGCACAGCGGAGCTGCGGTGTCAGTGGTACAGCGCGCA 2339
DB 2253 GTACGCTTTCAAATCAACCCCAAGCACAGCGGAGCTGCGGTGTCAGTGGTACAGCGCGCA 2312
QY 2340 TTGTCGCGCTTGGCGAAGACGATCCCGAGCTTTCAGCGGATCAAACTCTTGTACAGC 2399
DB 2313 TTGTCGCGCTTGGCGAAGACGATCCCGAGCTTTCAGCGGATCAAACTCTTGTACAGC 2372
QY 2400 ATTGCTTTGAGTACCGCGGCTGCTGCGCTGATTTTACCTAGGCGAGAGTGGGTACGCTC 2459
DB 2373 ATTGCTTTGAGTACCGCGGCTGCTGCGCTGATTTTACCTAGGCGAGAGTGGGTACGCTC 2432
QY 2460 AATGACGAGCTGCTGCTGCAAGA CAGCAATTAAGACGAGAGCAGCGCTTGGCGGAGCTG 2519
DB 2433 AATGACGAGCTGCTGCTGCAAG - CAGCAATTAAGACGAGAGCAGCGCTTGGCGG - CACCT 2490
QY 2520 CCGCGCTAACAGAGCGCTGTCGCAACGCAACGATCCGTCGACCGCAGCGCGGCA 2579
DB 2491 CCGCGCTAACAGAGCGCTGTCGCAACGCAACGATCCGTCGACCGCAGCGCGGCA -GGCA 2549
QY 2580 ATCTATCAGGCTTTCGCGCATATGATTTGCGTCCGCAAAAGCAATTCGCGCTTTCAGCGC 2639
DB 2550 ATCTATCAGGCTTTCGCGCATATGATTTGCGTCCGCAAAAGCAATTCGCGCTTTCAGCGC 2609
QY 2640 GCGAGGCTGTTTACATTTCAACCAACCAAGCAGCATCTCGGCTACATCCGCAACAT 2699
DB 2610 GCGAGGCTGTTTACATTTCAACCAACCAAGCAGCATCTCGGCTACAT -CGCAACAT 2668
QY 2700 GCGCTTTTGGCATTCGTTAACTTTCAGCGAATATCCGCAAAAGCTTTCAGCGGATACCTG 2759
DB 2669 GCGCTTTTGGCATTCGTTAACTTTCAGCGAATATCCGCAAAAGCTTTCAGCGGATACCTG 2728
QY 2760 CAGCGCATGCGCTTCAAGCGGACGACCTCATCGGTGGCAAACTGTCAGCGCTGAATCAG 2819

1663 AGCACCCGGGGGCTTCTCGCAACTGGAGAGCGGACGCTGGGTGFGAGACGACCTTCAATT 1722
1735 CTTCCCAATGGGACTTGAATTACAGCAACCCGCTGGTATTCGGGCGCAATGGCGCGGAGAA 1794
1723 CTTTCCAAATGGGACTTGAATTACAGCAACCCGCTGGTATTCGGGCGCAATGGCGCGGAGAA 1782
1795 TGTGTCTTCCCAATGGGCGGTGATCATCTCTGGTATGATGCGGTTCCTTTATTT 1854
1783 TGTGTCTTCCCAATGGGCGGTGATCATCTCTGGTATGATGCGGTTCCTTTATTT 1842
1855 GGAACCAATGGGAGCAAGCTTGGAAACCTTGGCGAGCGGCGGCTCATTCGGCGGT 1914
1843 GGAACCAATGGGAGCAAGCTTGGAAACCTTGGCGAGCGGCGGCTCATTCGGCGGT 1902
1915 TCAATGCCGTTATGGTATTCGGCGGCGGCGGCTTCTTCAAAATCGAAGCCATCGTCC 1974
1903 TCAATGCCGTTATGGTATTCGGCGGCGGCGGCTTCTTCAAAATCGAAGCCATCGTCC 1962
1975 ACCCGGACCAAGTCTCAATACATCGGCGAGGAGCAATGCCAATGGTTACACCCCC 2034
1963 ACCCGGACCAAGTCTCAATACATCGGCGAGGAGCAATGCCAATGGTTACACCCCC 2022
2035 TGCAAATGGCAATGTTGTGGAAACCTTGGCGGCGGAGTCAACCTGCTCCATCAGG 2094
2023 TGCAAATGGCAATGTTGTGGAAACCTTGGCGGCGGAGTCAACCTGCTCCATCAGG 2082
2095 CGTGACCTACCGCACAACTTGGCGGAGCATACCGGCTGGGTCACTACGTCGCGACCC 2154
2083 CGTGACCTACCGCACAACTTGGCGGAGCATACCGGCTGGGTCACTACGTCGCGACCC 2142
2155 ACGAGCATCGCTGGGAGCTTGGCGGAGTGAAGAGCGGCGCATATCTGGGCGATAGCGGCT 2214
2143 ACGAGCATCGCTGGGAGCTTGGCGGAGTGAAGAGCGGCGCATATCTGGGCGATAGCGGCT 2202
2215 ACGACCAACCGCAATTCCTCAACCGCTTCTTGGTCAACCGTTTGGACGCGAGCTTCGCTC 2274
2203 ACGACCAACCGCAATTCCTCAACCGCTTCTTGGTCAACCGTTTGGACGCGAGCTTCGCTC 2262
2275 GTGGGTACCGTTCCTCAATACACCCAGCGAGCGAGTGGCGTCTAGTGGTACAGCGG 2334
2263 GTGGGTACCGTTCCTCAATACACCCAGCGAGCGAGTGGCGTCTAGTGGTACAGCGG 2322
2335 CGGCATTTGCTGGCTTGGCGGAGAGCATTCGCCAGCGGTTGACGCGCATCAAACTCTTGT 2394
2323 CGGCATTTGCTGGCTTGGCGGAGAGCATTCGCCAGCGGTTGACGCGCATCAAACTCTTGT 2382
2395 ACAGCATTTGCTTGGTACCGGCTTGGCGGAGAGTTCCTAGTGGTACAGCGAGTGGGTA 2454
2383 ACAGCATTTGCTTGGTACCGGCTTGGCGGAGAGTTCCTAGTGGTACAGCGAGTGGGTA 2442
2455 CGCTCAATGACGACGACTGGTTCGCAAGCAGCAATAGAGCGAGCAGCGGTTGGGCGC 2514
2443 CGCTCAATGACGACGACTGGTTCGCAAGCAGCAATAGAGCGAGCAGCGGTTGGGCGC 2502
2515 ACCGTCCGCGCTTACACGAGGCTTGTAGCGGCAACGCAACGATCCGTGGACCGGAGCGG 2574
2503 ACCGTCCGCGCTTACACGAGGCTTGTAGCGGCAACGCAACGATCCGTGGACCGGAGCGG 2562
2575 GGCATATTCAGGCTTGGCGGATATGATTTGGCGTCCGCAAGAGCAATCGGCTTCG 2634
2563 GGCATATTCAGGCTTGGCGGATATGATTTGGCGTCCGCAAGAGCAATCGGCTTCG 2622
2635 ACGGCGGAGGCTTGTATCAATCAACCAACCAAGCAGCATTCGCTACATCCGCA 2694
2623 ACGGCGGAGGCTTGTATCAATCAACCAACCAAGCAGCATTCGCTACATCCGCA 2682
2695 ACAATGCGCTTGTGATTCGCTTACCTTACGAGCAATATCCGCAACCGTTACCGGCGATA 2754
2683 ACAATGCGCTTGTGATTCGCTTACCTTACGAGCAATATCCGCAACCGTTACCGGCGATA 2742
2755 CCTCGAAGCCATGCCCTTCAAGGCGGACGACTCATCGTGGCAAACTGTGACCTGA 2814

2743 CCTGCAAGCCATGCGCCTTCAAGGCGCAGCCTCATCTGCTGGCAAACTGTGAGCCTGA 2802
2815 ATCAGATTTGAGCGCTTCAGCCCTATCAGGTCTATGCTGCTCGAAATCGCTGACGACGC 2874
2803 ATCAGATTTGAGCGCTTCAGCCCTATCAGGTCTATGCTGCTCGAAATCGCTGACGACGC 2862
2875 TTC 2877
2863 CTC 2865
RESULT 7
AAF61710 ID AAF61710 standard; DNA; 1910 BP.
XX AAF61710;
XX 12-JUL-2001 (first entry)
XX Amylosucrase PCR derived DNA fragment AMSU5.
XX Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
XX poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
XX Unidentified.
XX WO200125449-A2.
XX 12-APR-2001.
XX 04-OCT-2000; 2000WO-EP09695.
XX 07-OCT-1999; 99DB-1048408.
XX (AXIV-) AXIVA GMEH.
XX Bengs H, Polakowski T, Held A, Gallert K;
XX MPI; 2001-328330/34.
XX Amylosucrase immobilized as fusion protein with anchoring sequence.
XX useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
Claim 3; Page 28; 38pp; German.
CC This invention describes a novel amylosucrase (AS), immobilizable on a
CC solid phase, which comprises a fusion protein (FP) of functional units
CC of AS, an anchoring sequence, and optionally additional auxiliary
CC sequences. The invention also describes (1) nucleic acid (I) encoding
CC FP; (2) expression vector containing (I) and able to express FP in a
CC host cell; (3) Escherichia coli containing the vector of (2); (3) anchoring
CC sequence, or its functional variants or fragments, of at least 8
CC nucleotides that encodes an epitope, a high-affinity binding partner or
CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS
CC comprising Glutathione-Sepharose; (5) combined, stable catalyst (A)
CC comprising FP immobilized on Sepharose for production of (II).
CC poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).
CC AS is used in production of poly(1,4-alpha-glucans) (II), useful for
CC producing films, as food additives, as starting materials for
CC cyclodextrins and as auxiliaries in pharmaceutical formulations.
CC Immobilized AS makes possible efficient, inexpensive and continuous
CC production of poly(1,4-alpha-glucans) (II), and it can be used
CC repeatedly. Compared with known methods, specificity is improved
CC (increased yield of (II) and reduced formation of palatinose) and
CC reaction is complete within 24 hours, compared to 48-72 hours for batch
CC methods. This sequence encodes a PCR derived amylosucrase (BC 2.4.1.4)
CC AMSU5 fragment described in the invention.
XX
SQ Sequence 1910 BP; 447 A; 595 C; 474 G; 387 T; 7 other;
Query Match 64.9%; Score 1891.4; DB 22; Length 1910;
Best Local Similarity 99.9%; Pred. No. 0;

Matches 1892; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	985	TGATTTTACAGTACCTCAAAAACACCGATCTTGGAATCTACACGCGCGGAACAGCGCGCG	1044
DB	11	TGATTTTACAGTACCTCAAAAACACCGATCTTGGAATCTACACGCGCGGAACAGCGCGCG	70
QY	1045	GCATCGAAAAATCCGAAAGACTGGCGCAGTTTTCGGCGCGCATGTGATACGCAATTTCCCCA	1104
DB	71	GCATCGAAAAATCCGAAAGACTGGCGCAGTTTTCGGCGCGCATGTGATACGCAATTTCCCCA	130
QY	1105	AACATGATGAACGAATCGACAGGGTGTACGGCAACAAACGAAGCCCTGCTGCTATGCTGG	1164
DB	131	AACATGATGAACGAATCGACAGGGTGTACGGCAACAAACGAAGCCCTGCTGCTATGCTGG	190
QY	1165	AAATGCTGCTGGCGGAGCATGCAAGAGCTATTTCCCAACGAACCTCATCTTTAAAGATA	1224
DB	191	AAATGCTGCTGGCGGAGCATGCAAGAGCTATTTCCCAACGAACCTCATCTTTAAAGATA	250
QY	1225	TGATATCGCGCGCGGAAACAAACCCCGATTGGATTGTTGTCACAAACAAAGTCGGCGGG	1284
DB	251	TGATATCGCGCGCGGAAACAAACCCCGATTGGATTGTTGTCACAAACAAAGTCGGCGGG	310
QY	1285	TGTCCTACGTTGATTTGTTGGCGGATTTGAGGGCTTGAAGATAAAATTCCTTAT	1344
DB	311	TGTCCTACGTTGATTTGTTGGCGGATTTGAGGGCTTGAAGATAAAATTCCTTAT	370
QY	1345	TTCAAGAGCTTGGTTGACTTATCTGCACCTGATGCCGCTGTTTAAATGCCCTGAAGCA	1404
DB	371	TTCAAGAGCTTGGTTGACTTATCTGCACCTGATGCCGCTGTTTAAATGCCCTGAAGCA	430
QY	1405	AAAGGACCGCGCTATGGGTGAGCAGCTACGGGATGTCATCCGGCAGCTGGGCACAA	1464
DB	431	AAAGGACCGCGCTATGGGTGAGCAGCTACGGGATGTCATCCGGCAGCTGGGCACAA	490
QY	1465	TAGCGACTTGGCGGAAGTCAATGCTGCGCTGCAAGACCGGCAATTTCCGCGCTGCTG	1524
DB	491	TAGCGACTTGGCGGAAGTCAATGCTGCGCTGCAAGACCGGCAATTTCCGCGCTGCTG	550
QY	1525	ATTATATCTTAACACACCTCTCAACGAAACAGATGGCGGACAGCTGGCGGCGGGG	1584
DB	551	ATTATATCTTAACACACCTCTCAACGAAACAGATGGCGGACAGCTGGCGGCGGGG	610
QY	1585	ACCGGCTTTTGACAAATTTCTACTATTTTCCCGACCGGATGTCGCGGCAATACG	1644
DB	611	ACCGGCTTTTGACAAATTTCTACTATTTTCCCGACCGGATGTCGCGGCAATACG	670
QY	1645	ACCGCACCTGGCGGAAATTTTCCCGACGAGCACCCGGCGGCTTCTCGAACTGGAAG	1704
DB	671	ACCGCACCTGGCGGAAATTTTCCCGACGAGCACCCGGCGGCTTCTCGAACTGGAAG	730
QY	1705	ACGGACGCTGGGTGGAGACCTTCAATTCCTTCCCAATGGGACTTGAATTAACAGCAAC	1764
DB	731	ACGGACGCTGGGTGGAGACCTTCAATTCCTTCCCAATGGGACTTGAATTAACAGCAAC	790
QY	1765	CGTGGGTATTTCCGCGCAATGGCGGGAATGCTGTTCTTGCCAACTTTGGCGGTGACA	1824
DB	791	CGTGGGTATTTCCGCGCAATGGCGGGAATGCTGTTCTTGCCAACTTTGGCGGTGACA	850
QY	1825	TCTCGGTATGATGGGTGGCTTCTTTATTTGGAAACAAATGGGCAAGCTGGGAACCC	1884
DB	851	TCTCGGTATGATGGGTGGCTTCTTTATTTGGAAACAAATGGGCAAGCTGGGAACCC	910
QY	1885	TGCGGAGGCGCACCGCCCTCATCCGCGGCTTCAATGCGGTTATGCGTATTTGCGGCGCG	1944
DB	911	TGCGGAGGCGCACCGCCCTCATCCGCGGCTTCAATGCGGTTATGCGTATTTGCGGCGCG	970
QY	1945	CGGTGTTCTTCAATCGAAGCATCGTCCACCCCGACCAAGTCGTCGAATACATCGGCG	2004
DB	971	CGGTGTTCTTCAATCGAAGCATCGTCCACCCCGACCAAGTCGTCGAATACATCGGCG	1030
QY	2005	AGGACGAATGCCAAATCGGTTACACCCCTGCAATGSCATTTGTGGAACACCCCTTG	2064
DB	1031	AGGACGAATGCCAAATCGGTTACACCCCTGCAATGSCATTTGTGGAACACCCCTTG	1090

RESULT 8

ID AAF61712 standard; DNA; 6851 BP.

XX AAF61712;

DT 12-JUL-2001 (first entry)

XX Expression vector pGEX-4T-1-AmSUS containing amylosucrase DNA.

XX Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;

KW poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.

OS Unidentified.

XX

QY	2065	CCACGCGGAAGTCAACTGCTCCATCAGGCGCTGACTACCGCCACAAACCTGCCGAGC	2124
DB	1091	CCACGCGGAAGTCAACTGCTCCATCAGGCGCTGACTACCGCCACAAACCTGCCGAGC	1150
QY	2125	ATACCGCTGGGTCAACTACGTCGCGAGCCACGACGACATCGGCTTGACGCTTGCAGTG	2184
DB	1151	ATACCGCTGGGTCAACTACGTCGCGAGCCACGACGACATCGGCTTGACGCTTGCAGTG	1210
QY	2185	AAGACGCGGCAATCTCGGCATTAAGCGCTACGACACCGCAATCTCTCAACCGCTTCT	2244
DB	1211	AAGACGCGGCAATCTCGGCATTAAGCGCTACGACACCGCAATCTCTCAACCGCTTCT	1270
QY	2245	TCGTCAACCCGTTTCGACGCGAGCTTCGCTCGTGGCGTACCGTTCCATATCAACCAAGCA	2304
DB	1271	TCGTCAACCCGTTTCGACGCGAGCTTCGCTCGTGGCGTACCGTTCCATATCAACCAAGCA	1330
QY	2305	CAGCGCATGCGGTGTCAGTGTGTACAGCGCGGATTTGGTGGCTTGGCGCAAGACGATC	2364
DB	1331	CAGCGCATGCGGTGTCAGTGTGTACAGCGCGGATTTGGTGGCTTGGCGCAAGACGATC	1390
QY	2365	CCACGCGGTTGACCGCATCAAACTCTTGTAACGCAATTTGCTTACGTCACCGCGCTGTC	2424
DB	1391	CCACGCGGTTGACCGCATCAAACTCTTGTAACGCAATTTGCTTACGTCACCGCGCTGTC	1450
QY	2425	CGTGAATTAACCTAGGCGACGAGTGGTACGCTCAATGACGACGACGCTGGTGGCAAGCA	2484
DB	1451	CGTGAATTAACCTAGGCGACGAGTGGTACGCTCAATGACGACGACGCTGGTGGCAAGCA	1510
QY	2485	GCAATTAAGACGACGACGACGCGTTCGGCGCACCGCTCCGCGCTCAACGGAAGCCCTGTACG	2544
DB	1511	GCAATTAAGACGACGACGACGCGTTCGGCGCACCGCTCCGCGCTCAACGGAAGCCCTGTACG	1570
QY	2545	CGCAACGCAACGATCCGTCGACCGGCGCGGCGGCAATCTATCAGGCTTGGCGCATATGA	2604
DB	1571	CGCAACGCAACGATCCGTCGACCGGCGCGGCGGCAATCTATCAGGCTTGGCGCATATGA	1630
QY	2605	TTGCGTCCGCGCAAGCAATCCGCGCTTCGACGCGGCGGAGCTGGTTACATTTCAACACCA	2664
DB	1631	TTGCGTCCGCGCAAGCAATCCGCGCTTCGACGCGGCGGAGCTGGTTACATTTCAACACCA	1690
QY	2665	ACAACAGCACATCATCGGCTACATFCGCGCAACATGCGCTTTTGGCATTCGCTTACCTCA	2724
DB	1691	ACAACAGCACATCATCGGCTACATFCGCGCAACATGCGCTTTTGGCATTCGCTTACCTCA	1750
QY	2725	GCGATATCCGCAACCGCTTACCGCGCATACCCCTGCAAGCCATGCCCTTCAAGGCGGACG	2784
DB	1751	GCGATATCCGCAACCGCTTACCGCGCATACCCCTGCAAGCCATGCCCTTCAAGGCGGACG	1810
QY	2785	ACCTCATCGGTGGCAAAACTGTCAAGCCCTGGAATCAGGATTTCAACGCTTCAGCCCTATCAGG	2844
DB	1811	ACCTCATCGGTGGCAAAACTGTCAAGCCCTGGAATCAGGATTTCAACGCTTCAGCCCTATCAGG	2877
QY	2845	TCATGTGCTGAAATCGCTGACGCGGCTTC	2877
DB	1871	TCATGTGCTGAAATCGCTGACGCGGCTTC	1903

XX New peptide encoded by operon including virulence genes of *Neisseria*
PT meningitidis, useful as vaccine component for treating or preventing
PT meningitis and for identifying antimicrobial drug -
XX
XX

AC	ABX03900;
XX	
DT	22-JAN-2003 (first entry)
XX	
DE	N. meningitidis DNA encoding a vaccine antigen #42.
XX	
KW	Vaccine; antigen; ds; meningococcal disease; pathogenic bacteria;
KW	meningitis.
XX	
OS	Neisseria meningitidis group B.
XX	
PN	WO200277648-A2.
XX	
PD	03-OCT-2002.
XX	
PF	22-MAR-2002; 2002WO-CB01399.
XX	
PR	22-MAR-2001; 2001GB-0007219.
XX	
PA	[MICR-] MICROBIOLOGICAL RES AUTHORITY.
PA	(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX	
PI	Robinson A, Gorringer AR, Hudson MJ, Bracegirdle P, West DM;
PI	Oliver KJ, Kroll JS, Langford PR;
XX	
DR	WPI; 2003-018958/01.
XX	
DR	P-PSDB; ABUC6063.
XX	
PT	Identifying an antigen for manufacturing a vaccine against
PT	meningococcal infection, comprises contacting antibodies with
PT	polypeptides, detecting polypeptide-antibody complexes, and identifying
PT	bound polypeptides as antigens -
XX	
PS	Claim 34; Page 239-240; 310pp; English.
XX	
CC	The invention relates to identifying an antigen comprising:
CC	(a) obtaining antibodies against a commensal bacteria, or an extract
CC	from a commensal bacteria; (b) contacting the antibodies with
CC	polypeptides obtained from an expression library of either a commensal or
CC	a pathogenic bacteria; (c) determining whether the polypeptides bind to
CC	antibodies; and (d) where a polypeptide binds to an antibody;
CC	identifying that polypeptide as an antigen. Also included are:
CC	(1) a method of preparing a vaccine composition, comprising identifying
CC	an antigen with the above method, and combining the antigen with a
CC	carrier; (2) a vaccine composition obtained by the above methods;
CC	(3) an antigen identified by the above methods; (4) a polypeptide
CC	encoded by all or a part of a nucleic acid sequence comprising the
CC	Neisseria lactamica DNA sequences detailed in the specification;
CC	(5) a vector comprising the nucleic acid molecule; (6) a method of
CC	preparing a composition for vaccination against infection by pathogenic
CC	bacteria, comprising: (a) obtaining a first antigen from a commensal
CC	Neisseria; (b) comparing the amino acid sequence of the first antigen
CC	with the amino acid sequence of the second antigen from a pathogenic
CC	bacteria, or comparing the sequence of a nucleic acid which codes for the
CC	first antigen with the sequence of the nucleic acid that codes for the
CC	second antigen; and if the first antigen is homologous to the second
CC	antigen or if the nucleic acid sequence for the first antigen is
CC	homologous to that of the second antigen, and (c) preparing a composition
CC	for vaccination against bacterial infection comprising the first antigen;
CC	and (7) an antibody that binds to the polypeptide antigen.
CC	The method is useful in screening commensal and pathogenic
CC	bacteria for previously unidentified vaccine antigens by identifying
CC	polypeptide antigens that bind to sera raised against commensal
CC	bacterial proteins. The polypeptide is useful as a vaccine antigen
CC	which may be used in the manufacture of a medicament for vaccination
CC	against meningococcal infection (e.g. meningitis). The present
CC	sequence encodes an antigenic protein from the pathogenic bacteria N.
CC	meningitidis.
XX	
XS	Sequence 765 BP; 167 A; 230 C; 216 G; 152 T; 0 other;

Query Match 21.9%; Score 639; DB 25; Length 765;
 Best Local Similarity 95.0%; Pred. No. 1,4e-165;

PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX (CHIR) CHIRON CORP.
 XX
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX WPI; 2000-318079/27.
 XX
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be
 PT used in the diagnosis and treatment of *N. meningitidis* infection and
 PT other *Neisseria* infections, for example, *N.gonorrhoea* -
 XX
 XX Claim 7; Page 1317-1331; 1760pp; English.
 XX
 XX The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AA81453 to AA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences; AA81260 to AA81303 and AA82562 to AA82563 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AA81254 to
 CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AA81322 to
 CC AA81452 represent *Neisseria meningitidis* MemB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament for in the manufacture of a
 CC medicament for treating, preventing or diagnosing infection due to
 CC *Neisseria* bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SQ Sequence 48275 BP; 10709 A; 12099 C; 13497 G; 11970 T; 0 other;
 Query Match 21.9%; Score 639; DB 21; Length 48275;
 Best Local Similarity 95.0%; Pred. No. 9 9e-165;
 Matches 660; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
 1 GAGTTTGGCTTCCGAAACGAAACGATGCTTTGAGCGGACACCTGTCCGCAAGCGG 60
 34649 GAGTTTGGCTTCCGAAACGAAACGATGCTTTGAGCGGACACCTGTCCGCAAGCGG 34708
 61 CTGACCGCCCTTTTGGCCATCGACATCGTAACATCGTTTGGTGGCAAGCTCTTTC 120
 34709 CTGACCGCCCTTTTGGCCATCGACATCGTAACATCGTTTGGTGGCAAGCTCTTTC 34769
 121 GCTTTGAGCGTGGCAGAAACAAAGTCAGACGCTCTTTCGCGCTTTGCGGCATCACCGCA 180
 34769 GCTTTGAGCGTGGCAGAAACAAAGTCAGACGCTCTTTCGCGCTTTGCGGCATCACCGCA 34828
 181 ATTTCAGATGTCGCGCGGAGTCTCCATCTGTTTCAGCGGATAGATTTCTTCT 240
 34829 ATTTCAGATGTCGCGCGGAGTCTCCATCTGTTTCAGCGGATAGATTTCTTCT 34889
 241 TGGCGGGCGTGGCGTGAACCTCATATTCAGACGAGGCGCGGATGCGGTTTTTGA 300
 34889 TGGCGGGCGTGGCGTGAACCTCATATTCAGACGAGGCGCGGATGCGGTTTTTGA 34948
 301 GATGCGCCACGCGCGCGGAGCGGCTTTTCGCGGAAACAAAGCTCGATATCGATAATG 360
 34949 GCATTTGGCCACGCGCGGAGCGGCTTTTCGCGGAAACAAAGCTCGATATCGATAATG 35008

QY 361 TGGGCGAGCGCGCTTTCAATCAGCGAGTGCAGCAGTTCAAAATAATAATCGTCCGAAAC 420
 DB 35009 TGGGCGAGCGCGCTTTTCGATCAGCGCGTGCAGCAGTTCAAAATAATAATCGTCCGAAAC 35068
 QY 421 GGGACAGCGCGCTTTTCGCGATGCGCTGAACTGAACGTAACGACGCGGCTTTCGCGGACG 480
 DB 35069 GGGACAGCGCGCTTTTCGCGATGCGCTGAACTGAACGTAACGACGCGGCTTTCGCGGACG 35128
 QY 481 CGCTCGCGAGCGCTTTCGCTGTGGCGCAATATCTTCGCGATGCTTCGCGCGCATTTCCAAA 540
 DB 35129 CGCTCGCGAGCGCTTTCGCTGTGGCGCAATATCTTCGCGATGCTTCGCGCGCATTTCCAAA 35188
 QY 541 AAATCGCGCGGAACTCGACATATCGAAGCGCAGTTTTCGATTTGCTGCAAGTACGCGG 600
 DB 35189 AAGTCGCGCGGAACTCGCAATATTCGAAAGGCGCATTTTTCGATTTGCTCAAGTACGCGG 35248
 QY 601 GAAAGTACGCGCGCATCGCGGCGACAGCGCGCATTTTGGTGGCTTCGCGCTTCG 660
 DB 35249 GAAAGTTCGCGCGCATCGCGGCGACAGCGCGCATTTTGGTGGCTTCGCGCTTCG 35308
 QY 661 ATAACGGTGTTCGACCGTTCAGCGTTCGCGTTCGAT 695
 DB 35309 ATAACGGTATTTTTCGACCAAGGCGGACGACAT 35343
 RESULT 12
 AAF21610/C
 ID AAF21610 standard; DNA; 349980 BP.
 XX
 AC AAF21610;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE *Neisseria meningitidis* B nucleotide sequence SEQ ID NO:111.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN WO2000066791-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05928.
 XX
 PR 30-APR-1999; 98US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 XX
 DR WPI; 2000-647603/62.
 XX
 PT *Neisseria meningitidis* B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent *Neisseria* infections -
 XX
 PS Claim 7; Appendix A; 692pp; English.
 XX
 CC The present invention describes the full length genome of
 CC *Neisseria meningitidis* B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the

CC Neisseria proteins given in AAB58550 to AAB58593, and AAB21589 to
CC AAB21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisseria bacteria or as a diagnostic reagent for detecting the
CC presence of Neisseria bacteria or of antibodies raised to Neisseria
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.

XX
SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;
Query Match 21.9%; Score 639; DB 21; Length 349980;
Best Local Similarity 95.0%; Pred. No. 2.5e-164;
Matches 660; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 GAGTTTGGCTCCCGACCGACGCTGATGCTTGAGCCGACACCTGTCCGGCAAGCGG 60
DB GAGTTTGGCTCCCGACCGACGCTGATGCTTGAGCCGACACCTGTCCGGCAAGCGG 290445

QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGGTTGGTGGCAAGCTCTTTC 120
DB CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGGTTGGTGGCAAGCTCTTTC 290385

QY 121 GCTTTGAGCGTGGCGAAGCAAGTCAAGCGTCTTCGCGCTTTGGGGATCACCGCA 180
DB GCTTTGAGCGTGGCGAAGCAAGTCAAGCGTCTTCGCGCTTTGGGGATCACCGCA 290325

QY 181 AFTTTGAGATGTCGCGCCGACGCTCTCTGTTTCAACGCGCATACGATTTCTTCT 240
DB AFTTTGAGATGTCGCGCCGACGCTCTCTGTTTCAACGCGCATACGATTTCTTCT 290265

QY 241 TCGCGCGCGTGGTGAACCTCATGATTCAGACGACGCGGCGGATGTCGCTTTTGA 300
DB TCGCGCGCGTGGTGAACCTCATGATTCAGACGACGCGGCGGATGTCGCTTTTGA 290205

QY 301 GCATGCGCCACGCGCGCGACGCGGTTTCGCGGGAAGAAAGCTCGATTCGATATG 360
DB GCATGCGCCACGCGCGCGACGCGGTTTCGCGGGAAGAAAGCTCGATTCGATATG 290145

QY 361 TCGCGCGCGCTTCAATCAGCGAGTCGACGAGTTCATAATATATCTCCGACAC 420
DB TCGCGCGCGCTTCAATCAGCGAGTCGACGAGTTCATAATATATCTCCGACAC 290085

QY 421 GGGAAACGAGCGCTTTCGCGGACGCTGAACTGAACTGAACTGAACTGAACTGAA 480
DB GGGAAACGAGCGCGCTTTCGCGGACGCTGAACTGAACTGAACTGAACTGAACTGAA 290085

QY 481 CGCTCGCGGACGCTGCTGCTGCGCAATCTTCGCGGATGCTGCGGCGATTCGCA 540
DB CGCTCGCGGACGCTGCTGCTGCGCAATCTTCGCGGATGCTGCGGCGATTCGCA 289965

QY 541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGTTTGTATTTGCTCAAGTACGCG 600
DB AAATCGCGCGGAACTCGACGATATCGAAGGCGAGTTTGTATTTGCTCAAGTACGCG 289905

QY 601 GAAAGTACGCGGATTCGCGGACGACGAGCGGCGGATTTGCTGCGTCCGCTTCG 660
DB GAAAGTACGCGGATTCGCGGACGACGAGCGGCGGATTTGCTGCGTCCGCTTCG 289845

QY 661 ATAACGCTGTTTGTGACCGCTCAGGCTGCTGTCAT 695
DB ATAACGCTGTTTGTGACCGCTCAGGCTGCTGTCAT 289810

RESULT 13
AAA81489/c
ID AAA81489 standard; DNA; 837096 BP.

XX AAA81489;
AC 04-DEC-2000 (first entry)
XX N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX Meningococcus B; MenB; ds.
XX Neisseria meningitidis.
XX Neisseria meningitidis.
XX WO200022430-A2.
XX 20-APR-2000.
XX 08-OCT-1999; 99WO-US23573.
XX 09-OCT-1998; 98US-0103794.
XX 30-APR-1999; 99US-0132068.
XX (CHIR) CHIRON CORP.
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
XX Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX Rappuoli R, Pizza M;
XX WPI; 2000-318079/27.
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
XX used in the diagnosis and treatment of N. meningitidis infection and
XX other Neisseria infections, for example, N.gonorrhoea -
XX Claim 7; Page 629-865; 1760pp; English.
XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX AAA8259 and AAA81304 to AAA81321 represent PCR primers used in the
XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies
XX against them, can be used in the manufacture of a composition. The
XX composition can be used as a medicament (or in the manufacture of a
XX medicament) for treating, preventing or diagnosing infection due to
XX Neisseria bacteria. For example, some of the identified proteins could
XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neisseriae. Identification of sequences
XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious
XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX Multivalent vaccines have also been tried but none have successfully
XX overcome antigenic variability. The provision of further, complete
XX sequences may provide an opportunity to identify secreted or surface
XX exposed proteins that may be presumed targets for the immune system and
XX which are not antigenically variable or at least more conserved than
XX other more variable regions.

XX
SQ Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;
Query Match 21.9%; Score 639; DB 21; Length 837096;
Best Local Similarity 95.0%; Pred. No. 3.9e-164;
Matches 660; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 GAGTTTGGCTCCCGACCGACGCTGATGCTTGAGCCGACACCTGTCCGGCAAGCGG 60
DB 55275 GAGTTTGGCTCCCGACCGACGCTGATGCTTGAGCCGACACCTGTCCGGCAAGCGG 55216

QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGTGTGCTGCGCAAGCTCTTTC 120

Db 55215 CTGACCGCCCGCTCTGCCCATCGACATCGTAACAATCGGTTGCGGCAAGCTCTTTC 55156
Qy 121 GCTTTGAGGCTGCGCAGAAAGTACAGCATCTTTCGCGCTTTGCGGCATCACGCA 180
Db 55155 GCTTTGAGGCTGCGGAAAGTACAGCATCTTTCGCGCTTTGCGGCATCACGCA 55096
Qy 181 ATTTTCAGATGTCGCGGCGAGTCTCCATCTGTTTCAGAGCGCATAGATTCTTCT 240
Db 55095 ATTTTCAGATGTCGCGGCGAGTCTCCATCTGTTTCAGAGCGCATAGATTCTTCT 55036
Qy 241 TCGCGCGGCTGCGGTGAATCTATGATTCAGAGCGCGCGCATGCGGTTTTTTTGA 300
Db 55035 TCGCGCGGCTGCGGTGAATCTATGATTCAGAGCGCGCGCATGCGGTTTTTTTGA 54976
Qy 301 GCATGCGCAGCGCGCGCGAGCGGCTTTTCGCGGAAAAAGCTCGATATCGATAATG 360
Db 54975 GCATTTGCCACGCGCAGCGCGAGCGGCTTTTCGCGGAAAAAGCTCGATATCGATAATG 54916
Qy 361 TCGCGCAGCGGCTTTCATGATCGAGCGCTCGAGCATTTCAAAATATAATCGTCGACAC 420
Db 54915 TCGCGCAGCGGCTTTCATGATCGAGCGCTCGAGCATTTCAAAATATAATCGTCGACAC 54856
Qy 421 GGAACAGCGCGCTTCGCGCATCGCTGTAACGTAACGTAACGAGCGGCTTTCGCGCAGC 480
Db 54855 GGAACAGCGCGCTTCGCGCATCGCTGTAACGTAACGAGCGGCTTTCGCGCAGC 54796
Qy 481 GCGTCGCGGAGCGTCTCGGTCGCGCAATCTTTCGCGCATGTCGCGCGCATTCGAA 540
Db 54795 GCGTCGCGGAGCGTCTCGGTCGCGCAATCTTTCGCGCATGTCGCGCGCATTCGAA 54736
Qy 541 AAATCGCGCGGAACTCGAGATATCGAAGCGGAGGTTTTTGAATTTGGTCAAGTACGCG 600
Db 54735 AAGTCGCGCGGAACTCGGCAATATCGAAGCGCATATTTTGAATTTGCTCAAGTACGCG 54676
Qy 601 GAAATACGCGCGCATCGCGGCGAGCGAGCGCGGATTTTGGTTCGCTCGCTTCGCG 660
Db 54675 GAAATTCGCGCGCATCGCGGCGAGCGAGCGCGGATTTTGGTTCGCTTCGCGCTTCG 54616
Qy 661 ATAACGGTGTTTTGAAGCTGAGCTGTGTGCAT 695
Db 54615 ATAACGGTATTTTGAACAAGCGAGGACACAT 54581

RESULT 14

ABZ40085/c

ID ABZ40085 standard; DNA; 762 BP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a
PT medicament for treating or preventing *N. gonorrhoeae* infection -
XX
PS Disclosure; Page 536; 815pp; English.
XX
CC The present invention relates to proteins from *Neisseria gonorrhoeae*.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC molecules of the invention.

XX Sequence 762 BP; 161 A; 238 C; 214 G; 149 T; 0 other;

Query Match 21.9%; Score 637.4; DB 25; Length 762;

Best Local Similarity 94.8%; Pred. No. 3.7e-165;

Matches 659; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1 GAGTTTTCGTTCCCGAACCGGAACTGATGCTTTCGAGCGGAAACACCTGTCGCGCAAGCGG 60

Db 695 GAGTTTTCGTTCCCGAACCGGAACTGATGCTTTCGAGCGGAAACACCTGTCGCGCAAGCGG 636

Qy 61 CTGACCGCGCCCTTTTTCGCGCATCGACATCGTAAACATCGGTTTTCGCGCAAGCTTCTTTC 120

Db 635 CTGACCGCGCGCTTTCGCGCATCGACATCGTAAACATCGGTTTTCGCGCAAGCTTCTTTC 576

Qy 121 GCTTTTCGAGCGTTCGAGAAAGCAAGTTCAGCGCTTTCGCGCTTTCGCGCAAGCTTCTTTC 180

Db 575 GCTTTTCGAGCGTTCGAGAAAGCAAGTTCAGCGCTTTCGCGCTTTCGCGCAAGCTTCTTTC 516

Qy 181 ATTTTCGAGATGTCGCGCGCGAGTCTTTCATCTGTTTCGAGCGGATACGATTTCTTCT 240

Db 515 ATTTTCGAGATGTCGCGCGCGAGTCTTTCATCTGTTTCGAGCGGATACGATTTCTTCT 456

Qy 241 TCGCGCGCGCTTCGAGTAACTCATGATTCGAGCGGCGCGGATGTCGCGCTTTCGCGCAAG 300

Db 455 TCGCGCGCGCTTCGAGTAACTCATGATTCGAGCGGCGCGGATGTCGCGCTTTCGCGCAAG 396

Qy 301 GCATGCGCCACCGCGCGCGGAGTCTTTCGCGGAAAAAGCTTCGATATCGATAATG 360

Db 395 GCATTTGCCACCGCGCGCGGAGTCTTTCGCGGAAAAAGCTTCGATATCGATAATG 336

Qy 361 TCGCGCGCGCTTTCATGAGCGGATGAGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 420

Db 335 TCGCGCGCGCTTTCATGAGCGGATGAGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 276

Qy 421 GCGAACGAGCGCGCTTTCGCGCATGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 480

Db 275 GCGAACGAGCGCGCTTTCGCGCATGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 216

Qy 481 GCGTCGCGGAGCGCTTTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 540

Db 215 GCGTCGCGGAGCGCTTTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 156

Qy 541 AAATCGCGCGGAACTTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 600

Db 155 AAGTCGCGCGGAACTTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 96

Qy 601 GAAATACGCGGCGATTCGCGGCGGAGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 660

Db 95 GATTAATTCGCGGCGATTCGCGGCGGAGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 36

Qy 661 ATAACGGTGTTCGAGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 695

Db 35 ATAACGGTATTTTTCGAGCGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAGGATTCGAG 1

RESULT 15

AAS59546/c

ID AAS59546 standard; DNA; 29559 BP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

Fontana MR, Pizza M, Masignani V, Monaci E;

WPI; 2003-058415/05.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

P-PSDB; ABP75115.

AC AAS59546;
XX 13-FEB-2002 (first entry)
XX Propionibacterium acnes immunogenic protein encoding DNA #41.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.
XX
OS Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
PI
XX WPI; 2001-616774/71.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Claim 1; SEQ ID No 41; 1069pp; English.
XX
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by P. acnes. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis, and endophthalmitis. P. acnes is also involved
CC in infections of bone, joints and the central nervous system, however it
CC is particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acnes in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for P. acnes proteins. These antibodies can be
CC used to downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AAU49884-AAU50191 and AAU67524-AAU67527.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 29559 BP; 5419 A; 8656 C; 9368 G; 6114 T; 2 other;
Query Match 10.2%; Score 297.2; DB 23; Length 29559;
Best Local Similarity 52.8%; Pred. No. 1.3e-70;
Matches 790; Conservative 0; Mismatches 658; Indels 48; Gaps 5;
QY 1189 AAAGCTATTCCCAACGCACTCATCTCTAAAGATATCGATTCGGCGGCGAAACAACC 1248
DB 21955 AACGGTGTGCGGACGCTCTGATGATCTCAAGCGCTCGAAGCCGCTCTCTGGAGC 21896
QY 1249 CGATTGGATTTTGGCCAAACAAAGTCGGGGGGGTGCTACGTGATTTGTTGGCG 1308
DB 21895 CCGACTGGCTTCAACAAACGACGATGATCGTTATGCAACCTTATACCGACCATTTCTCG 21836
QY 1309 GCGATTGAAGGGCTTGAAGATAAAATTCCTTAATTTCAAGAGCTTGGTTGACTATTC 1368

DB 21835 GCACCTCAAAGGTATAAGCGATCACCTCGATCACTCTCGCATGGGTGTCGCTATC 21776
QY 1369 TGCACTGATGCGCTGTTTAAATGCTTGAAGCAAAAGCGAGCGGCTATCGGTCGA 1428
DB 21775 TGCATCTCATGCCCTGCTGCAACCTCGCCAGGTACTGACGATGGTGTATCGGTG 21716
QY 1429 GCAGTACCGGATGTCATTCGCGACTGGGCAATAGGAGCTTGGCGGAAGTCAATG 1488
DB 21715 CGGACCATCGCACTATCCGCAOAGACTGCGGTACTATGAGTACCTAGCTGACCTACG 21556
QY 1489 CTGCGCTGCACGAAGCGCGCATTTCCGCGCTGTCGATTTTATCTTCAACACCACTCCA 1548
DB 21555 CTACCTGCGTGCCCAAGCATATCTCTCATGAGCTCATGTTAACCACTAGCGG 21596
QY 1549 ACGAACAGAAATGGGCGCAACGCTGCGCGCGCGGAGCCGCTTTTCGACAAATTTCTACT 1608
DB 21595 CCGAGCATGAATGGGCTCGAAGAGCTCGGCTGCTCAACAGAGTACCGGACTATTTCC 21536
QY 1609 ATATTTTCCCGGACGCGGATGCCGACCAATACGACCGCACCTCGCGGAAATCTTCC 1668
DB 21535 ATATCTGTGACTCAAGATGAGGTTGACGCTGGGAGAGAACCTTCGGAAGTATTTTC 21476
QY 1669 CCGACAGCACCGCGGCTTTCTGCAACTGGAAGACCGAC---GCTGGGTGFGAGCA 1725
DB 21475 CGGACTTCGCGCATGGCAATTTCACTGGGACGACGACTGCGAGGGTGGTATGGCGCA 21416
QY 1726 CTTCAATTCCTCCATGGGACTTGAATTACAGCAACCGTGGTATTTCGCGCATGG 1785
DB 21415 CTTTCAAGAAATCCAGTGGGATCTCACTGGGCTAACTGAGCTTTCTGCAATTC 21356
QY 1786 CGGCGGAAATGCTGTTCTTTCGCAACTTGGCGGTGACATCTCTGCTATGATGCGGTG 1845
DB 21355 TCGACCTGATGGGTGCTGCTGCCAATCGCGGGTTCGAGGTGTTCCGCTCGAGCCATCG 21296
QY 1846 CTTTATTTGGAACAAATGGGCAAGCTGCGAAACCTGCGGACGCGCAGCTCTCA 1905
DB 21295 CTTTATCTGGAAGAACTCGGCACTGTCAGAACTTTCGGAATTTCCAGCAATCA 21236
QY 1906 TCGCGCGTTCAATGCGGTTATGCGCGCGCGCGCTGTTCTTCAAAATCCGAAG 1965
DB 21235 CTGAGTCATTCGCTCAGGCAATACGATTCGCGCGCGCGCTGTTCTTTCATGCGCGATG 21176
QY 1966 CCAATGTCACCGCGCAAGAGTGTCCAAATACATTCGCGGAGG-----ACG 2010
DB 21175 CCAATGTTGGCGCGAGTCTACGCGGATATTTCCGAGCGCGGCGGCGGCTGCGGAGAAAG 21116
QY 2011 AATGCCAATTCGGTTACACCCCTGCAATGCAATGTTGTGGAACACCTTCGCCACGC 2070
DB 21115 TCTCGGACATGATCTATCAACAGCTCATGTTGCGAGCTGTGAGCGCGCTAGTACCC 21056
QY 2071 GCGAAGTCAACCTGCTCATCAGGCGCTGACCTACCGCCCAACCTTCGCGAGCATACCG 2130
DB 21055 GCGAGCTCAGCTCATGGAACACAGCGTTGAGTCGAGCGCGCGCAACACCTTCGACACAA 20996
QY 2131 CTTGGGTCACTAGCTCGGACGCGACGACGACATCGGCTGGAGCTTTGCGGATGAGAGC 2190
DB 20995 CTTGGGCGACCTACGCTCGATGTCACGAGCATCGGATGCGATGATGATGCGGAGC 20936
QY 2191 CGGCATATCTGGGCAATAAGCGGCTTACGACCAACCGCCCAATTCCTCAACGCTTCTTCTGCTCA 2250
DB 20935 CCGGTAAACCTGGCTTGAACCCGCTGCGCCACGCGAGTTCTCTCGACTTCTTATTCG 20876
QY 2251 ACCGTTTCGAGCGGAGCTTGGCTGTCGCGTACCGTTCCATATACACCAAGCAGCAGCG 2310
DB 20875 GGACATTCCTCGGATCTCTCGCGCTGCTGCTGCTTTCGAAGACACCCGCTTACCGGCG 20816
QY 2311 ACTGCGGTGTCAGTGTACAGCGCGGCTATGGCTGGCTGG-----2352
DB 20815 ACCGTGAATTTAGTGGCTCTTACAGAGCTGGCGGGCTGGAAATCTTCCCTCGAGTCCG 20756
QY 2353 -----CGCAAGACGATCCCCACGCGGTGTGACCGCATCAAACTCTTGTATAGAGCATTTGCTT 2406

```
Db      20755  ACGACCCAGCTGGTGTGAGCGTGCATCGCCCGGATCGTTATGCTGCACACGGCGATCC 20696
Qy      2407  TGAGTACCGCGCGTCTGCCCGTGAATTA CCTAGCGGACGAGTGGGTACGCTCAATGACG 2466
Db      20695  TCGGGTACGCGCGAGTACCCTAACTCTGGATGGGACGAGGTTCGGATGCTCA---ACG 20639
Qy      2467  ACGACTGGTCGCAAGACAGCAATATACAGCGACGACGCGTGGGGCGACCGTCCGGCGCT 2526
Db      20638  ACGACTGGCAACGTGATCCCGGTCTATGCCGACGACACCGCTGGGTCCATCGACCTATGA 20579
Qy      2527  ACAACGAAGCCCTGTATCGGCGCA---ACGCAACGATCCGTGACCGCGAGCCGGGCAAACT 2583
Db      20578  TGAACCTGGTCGATGGTCAAGCAGGCCCA CGCCGAGCCGACAGCGTTCTGGTGAATAT 20519
Qy      2584  ATCAGGGCTTGGCCATATGATGCCGTGGCCAAAGCAATCGCGCTTCGAAGGC 2639
Db      20518  GGAACGGGTGCGCCCGGCCATCAATGCCCCGACACCGCGAGCCCGAGTTCCATGCC 20463
```

Search completed: November 8, 2003, 13:08:18
Job time : 743 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 13:08:21 ; Search time 193 Seconds
(without alignments)
6664.199 Million cell updates/sec

Title: US-09-843-007A-1
Perfect score: 2914
Sequence: 1 GAGTTTGGTTCCGACACC.....CGTTTCAGCGCATTTGCG 2914

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A COMB seq: *
2: /cgn2_6/prodata/2/ina/5B COMB seq: *
3: /cgn2_6/prodata/2/ina/6A COMB seq: *
4: /cgn2_6/prodata/2/ina/6B COMB seq: *
5: /cgn2_6/prodata/2/ina/6C COMB seq: *
6: /cgn2_6/prodata/2/ina/6D COMB seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2914	100.0	2914	3	US-08-737-752A-1
2	125.2	4.3	3414	4	US-09-252-991A-8258
3	125.2	4.3	4125	4	US-09-252-991A-8065
4	104.4	3.6	1704	1	US-08-528-199-2
5	104.4	3.6	1704	1	US-08-528-199-5
6	99.4	3.4	4403765	3	US-09-103-840A-2
7	99.4	3.4	4411529	3	US-09-103-840A-1
8	86.6	3.0	768	4	US-09-107-532A-2141
9	84.6	2.9	873	4	US-09-328-352-3264
10	73.6	2.5	2889	1	US-08-537-002A-4
11	73.6	2.5	2889	3	US-08-863-010-4
12	73.6	2.5	2889	3	US-09-024-402A-4
13	73.6	2.5	3600	1	US-08-537-002A-5
14	73.6	2.5	3600	3	US-08-863-010-5
15	73.6	2.5	3600	3	US-09-024-429-5
16	59.8	2.1	1782	1	US-08-374-155A-13
17	59.8	2.1	1782	2	US-08-785-396-13
18	58.6	2.0	753	4	US-09-134-001C-2602
19	58.6	2.0	4776	2	US-08-852-401-1
20	57.6	2.0	1728	4	US-09-107-532A-2079
21	49.4	1.7	1635	4	US-09-107-532A-482
22	47	1.6	4403765	3	US-09-103-840A-2
23	47	1.6	4411529	3	US-09-103-840A-1
24	46.8	1.6	1704	1	US-08-374-155A-15
25	46.8	1.6	1704	1	US-08-785-396-15
26	46.4	1.6	1509	4	US-09-252-991A-11048
27	46.4	1.6	1590	4	US-09-252-991A-11206

C 28 45.8 1.6 4105 4 US-09-634-238-182
29 45.8 1.6 4784 4 US-09-634-238-24
30 43.4 1.5 1842 4 US-09-252-991A-5853
31 43.4 1.5 2034 4 US-09-252-991A-5871
32 43.4 1.5 6142 4 US-09-514-302-1
33 42.8 1.5 1488 4 US-09-252-991A-5798
34 42.4 1.5 77536 4 US-09-410-551B-1
35 41.4 1.4 1872 3 US-08-818-111-17
36 41.4 1.4 1872 4 US-09-056-556-17
37 41.4 1.4 1872 4 US-09-072-596-17
38 41.4 1.4 1650 4 US-09-107-532A-3186
39 41 1.4 2244 1 US-08-476-519-10
40 41 1.4 2244 5 PCT-US95-09323-10
41 41 1.4 2334 1 US-08-476-519-1
42 41 1.4 2334 5 PCT-US95-09323-1
43 41 1.4 9223 4 US-08-961-527-59
44 40.8 1.4 9223 4 US-09-410-551B-1
C 45 40 1.4 77536 4 US-09-410-551B-1

ALIGNMENTS

RESULT 1
US-08-737-752A-1
; Sequence 1, Application US/08737752A
; Patent No. 6265635
; GENERAL INFORMATION:
; APPLICANT: Rossmann, Jens
; APPLICANT: Buttcher, Volker
; APPLICANT: Welsh, Thomas
; TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE
; OF FACILITATING THE SYNTHESIS OF LINEAR
; TITLE OF INVENTION: ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND
; MICROORGANISMS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: James F. Haley, Jr., c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,752A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 17 879.4
FILING DATE: 18-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 47 388.5
FILING DATE: 22-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: James F. Haley, Jr., Esq.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: GFB-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2914 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Neisseria polysacchara
IMMEDIATE SOURCE:
LIBRARY: genomic library in pBluescriptII SK
CLONE: pNB2
FEATURE:
NAME/KEY: CDS
LOCATION: 957..2867
US-08-737-752A-1

Query Match      100.0%; Score 2914; DB 3; Length 2914;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2914; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGTTTGGTTCCTCCGAAACCGAAGCGATGCTGAGCCGACCACTGTCGCGCAAGCGG 60
DB 1 GAGTTTGGTTCCTCCGAAACCGAAGCGATGCTGAGCCGACCACTGTCGCGCAAGCGG 60
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACAATCGTTTGGTGGCAAGCTCTTTTC 120
DB 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACAATCGTTTGGTGGCAAGCTCTTTTC 120
QY 121 GCTTTGAGCGTGGCGAAGAAAGTACGACGCTCTTCCGCGCTTTGGCGCATCAGCGCA 180
DB 121 GCTTTGAGCGTGGCGAAGAAAGTACGACGCTCTTCCGCGCTTTGGCGCATCAGCGCA 180
QY 181 ATTTTCCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAAGCGCATACGATTTCTTCT 240
DB 181 ATTTTCCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAAGCGCATACGATTTCTTCT 240
QY 241 TCGCGCGCGTGGCGGTGAACTCATGATTCAGAGAGCGCGCGCATGCGGTTTTTTCGA 300
DB 241 TCGCGCGCGTGGCGGTGAACTCATGATTCAGAGAGCGCGCGCATGCGGTTTTTTCGA 300
QY 301 GCATGCGCAGCGCGCGCGAGCGGCTTCCCGGAAAGAGCTCGATATCGATAATG 360
DB 301 GCATGCGCAGCGCGCGCGAGCGGCTTCCCGGAAAGAGCTCGATATCGATAATG 360
QY 361 TCGCGCAGCGCGCTTCAATCAGCGATCGAGAGTTCAAAATTAATGTCGCAACAC 420
DB 361 TCGCGCAGCGCGCTTCAATCAGCGATCGAGAGTTCAAAATTAATGTCGCAACAC 420
QY 421 GGGACAGCGCGCTTCCGATCGCGCTGAAAGTAAACAGCAGCGCTTGTGCGGAGC 480
DB 421 GGGACAGCGCGCTTCCGATCGCGCTGAAAGTAAACAGCAGCGCTTGTGCGGAGC 480
QY 481 GCGTCGCGGACGCTCTCGGTGTGCGCAATCTTCGCGATGCTGCCCGCGCATTCGAA 540
DB 481 GCGTCGCGGACGCTCTCGGTGTGCGCAATCTTCGCGATGCTGCCCGCGCATTCGAA 540
QY 541 AAATCGCGCGGAATCGACGATATCGAAGGCAAGGTTTGTGATTTGGTCAAGTACGGG 600
DB 541 AAATCGCGCGGAATCGACGATATCGAAGGCAAGGTTTGTGATTTGGTCAAGTACGGG 600
QY 601 GAAGTACGCGGCGCATCGCGGCGAAGCGGACGCGGATTTGGTGGTTCGCTTCCG 660
DB 601 GAAGTACGCGGCGCATCGCGGCGAAGCGGACGCGGATTTGGTGGTTCGCTTCCG 660
QY 661 ATAAAGGTTTGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 661 ATAAAGGTTTGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 721 GGTAAAGCGCAATATAGCAAGGCGCAGCGAATGTTTCAGACGCGATTTCTGTGGGC 780
DB 721 GGTAAAGCGCAATATAGCAAGGCGCAGCGAATGTTTCAGACGCGATTTCTGTGGGC 780
QY 781 CGGCTTGATATCAATCAAGCAGCATTCGCAATATCGAATGCGATTTGGCAAGCCCTG 840
DB 781 CGGCTTGATATCAATCAAGCAGCATTCGCAATATCGAATGCGATTTGGCAAGCCCTG 840
QY 841 TCCTTTCTAGTCAGTCGCGAGTCTTCTGAGTATGATTCGACGACCGCCCTACAGGCAT 900
DB 841 TCCTTTCTAGTCAGTCGCGAGTCTTCTGAGTATGATTCGACGACCGCCCTACAGGCAT 900
```

```
DB 841 TCCTTTCTAGTCAGTCGCGAGTCTTCTGAGTATGATTCGACGACCGCCCTACAGGCAT 900
QY 901 TTGCAAGATACGGCGGAGAGCGCGCGTTCGAAATCTCAGAACTCGAGCAGGATCATGT 960
DB 901 TTGCAAGATACGGCGGAGAGCGCGCGTTCGAAATCTCAGAACTCGAGCAGGATCATGT 960
QY 961 TGACCCCGACGACGAAAGTGGTTTGAATTTACAGTACTCAGAACTCAAAACAGCATCTTGGACA 1020
DB 961 TGACCCCGACGACGAAAGTGGTTTGAATTTACAGTACTCAGAACTCAAAACAGCATCTTGGACA 1020
QY 1021 TCTACAGCGCGAAGACGCGCGCGCATCGAAATCGAAGACTGGCGGCGAGTTTTCG 1080
DB 1021 TCTACAGCGCGAAGACGCGCGCGCATCGAAATCGAAGACTGGCGGCGAGTTTTCG 1080
QY 1081 GCGCATGAGATACGCAATTTCCCAACTGATGAAGAACTCGACAGACGCTGTACGCAACA 1140
DB 1081 GCGCATGAGATACGCAATTTCCCAACTGATGAAGAACTCGACAGACGCTGTACGCAACA 1140
QY 1141 ACGAAGCCCTGCTGCTATGCTGGAATGCTGCTGGCGCAGGCAATGCAAAAGCTATTC 1200
DB 1141 ACGAAGCCCTGCTGCTATGCTGGAATGCTGCTGGCGCAGGCAATGCAAAAGCTATTC 1200
QY 1201 AACGCAACTCATCTTAAAGATATCGATATCGCGCGGCAAAACAAACCCGATTCGATTT 1260
DB 1201 AACGCAACTCATCTTAAAGATATCGATATCGCGCGGCAAAACAAACCCGATTCGATTT 1260
QY 1261 TGTCCAAACAAAGTTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1261 TGTCCAAACAAAGTTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 1321 GCTTCAAGATATGAAATTTCTTATTTTCAAGAGCTTTGGTTTGAATTTATCTGCACTTAT 1380
DB 1321 GCTTCAAGATATGAAATTTCTTATTTTCAAGAGCTTTGGTTTGAATTTATCTGCACTTAT 1380
QY 1381 CGCTGTTAAATGCTTGAAGCAAGGAGAGCGCGGCTATGCGGTGAGCAGCTACCGG 1440
DB 1381 CGCTGTTAAATGCTTGAAGCAAGGAGAGCGCGGCTATGCGGTGAGCAGCTACCGG 1440
QY 1441 ATGTCAATTCGCGCATCTGGGCAATAGCGCATCTTCGCGAATCTATGCTGCTGCTGCTG 1500
DB 1441 ATGTCAATTCGCGCATCTGGGCAATAGCGCATCTTCGCGAATCTATGCTGCTGCTGCTG 1500
QY 1501 AAGCGCGCATTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
DB 1501 AAGCGCGCATTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
QY 1561 GGGCGCGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620
DB 1561 GGGCGCGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620
QY 1621 ACCGCGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
DB 1621 ACCGCGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
QY 1681 CGGCGCGCTTCTGCGCACTGGAGAGCGGAGCGGCTGCGGTGAGCAGCTTCAATTCCTTCC 1740
DB 1681 CGGCGCGCTTCTGCGCACTGGAGAGCGGAGCGGCTGCGGTGAGCAGCTTCAATTCCTTCC 1740
QY 1741 AATGGAGCTTGAATTAACAGCAACCGCGGTATTCGCGCAATGCGCGCGCAATGCGGTGT 1800
DB 1741 AATGGAGCTTGAATTAACAGCAACCGCGGTATTCGCGCAATGCGCGCGCAATGCGGTGT 1800
QY 1801 TCTTCCCACTTGGCGGTGACATCTCTGCTATGAGTGGGTGCTGCTTATTTTGGAAAC 1860
DB 1801 TCTTCCCACTTGGCGGTGACATCTCTGCTATGAGTGGGTGCTGCTTATTTTGGAAAC 1860
QY 1861 AAATGGGAGAGCTGCGGAAACCTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
DB 1861 AAATGGGAGAGCTGCGGAAACCTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
QY 1921 CGGTTATGCGTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
DB 1921 CGGTTATGCGTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
```

1981 ACCAAGTGTGTCATATACATCGGCGCAGGACGAATGCAATCGTTTACACCCCTGCAAA 2040
1981 ACCAAGTGTGTCATATACATCGGCGCAGGACGAATGCAATCGTTTACACCCCTGCAAA 2040
2041 TGGCAATGTTGGAACACCTTGGCAACCGCGAGAGTCAACTGCTCCATCAGCGCTGA 2100
2041 TGGCAATGTTGGAACACCTTGGCAACCGCGAGAGTCAACTGCTCCATCAGCGCTGA 2100
2101 CCTACCGGCACAACTGCGCGAGCATACCGCTGCTGCTCAACTGCTCCGACGACGAG 2160
2101 CCTACCGGCACAACTGCGCGAGCATACCGCTGCTGCTCAACTGCTCCGACGACGAG 2160
2161 ACATCGGCTGGAAGTGTGCGCGATGAGAGCGCGGCAATCTGGGCAATAGCGGCTACGAC 2220
2161 ACATCGGCTGGAAGTGTGCGCGATGAGAGCGCGGCAATCTGGGCAATAGCGGCTACGAC 2220
2221 ACCGCAATCTCTCAACCGCTTCTGCTCAACCGCTTTCGACGCGAGCTTGGCTGCGG 2280
2221 ACCGCAATCTCTCAACCGCTTCTGCTCAACCGCTTTCGACGCGAGCTTGGCTGCGG 2280
2281 TACCGTTCGAATACCAACCAAGCAGAGCGAGTGGCTGCTGAGTGACAGCGCGGAT 2340
2281 TACCGTTCGAATACCAACCAAGCAGAGCGAGTGGCTGCTGAGTGACAGCGCGGAT 2340
2341 TGGTGGCTTGGCGCAAGACGATCCCGCGCTGACCGCATCAAACTCTTTGTACAGCA 2400
2341 TGGTGGCTTGGCGCAAGACGATCCCGCGCTGACCGCATCAAACTCTTTGTACAGCA 2400
2401 TTGCTTTGAGTACCGCGCTGCTGCTGCTGATTTACTAGCGCAAGTGGGTACGCTCA 2460
2401 TTGCTTTGAGTACCGCGCTGCTGCTGCTGATTTACTAGCGCAAGTGGGTACGCTCA 2460
2461 ATGACGACGCTGCTGCAAGCAGCAGTAAAGCAGCAGCAGCGCTTGGGCGCACCGTC 2520
2461 ATGACGACGCTGCTGCAAGCAGCAGTAAAGCAGCAGCAGCGCTTGGGCGCACCGTC 2520
2521 CCGCTTACCAACGAAGCCCTGTACGCGCAACGACGATCCGTCGACCGCGCGGCAAA 2580
2521 CCGCTTACCAACGAAGCCCTGTACGCGCAACGACGATCCGTCGACCGCGGCAAA 2580
2581 TCTATCAGGGCTTGGCGCATATGATTTGCGCTGCGCAAGAGATCGCGCTTCAAGCGG 2640
2581 TCTATCAGGGCTTGGCGCATATGATTTGCGCTGCGCAAGAGATCGCGCTTCAAGCGG 2640
2641 GCAGGTGCTTACATTCACACCAACCAAGCAGATATCGGCTACATCCGCAACATG 2700
2641 GCAGGTGCTTACATTCACACCAACCAAGCAGATATCGGCTACATCCGCAACATG 2700
2701 CGCTTTGCAATTCGCTAACTTCAAGCAATATCGGCAACCGTTACCGCGCATACCTG 2760
2701 CGCTTTGCAATTCGCTAACTTCAAGCAATATCGGCAACCGTTACCGCGCATACCTG 2760
2761 AAGCCATGCGCTTCAAGCGCAGCCTCATCGGTGGCAAACTGTGACGCTGATTCAG 2820
2761 AAGCCATGCGCTTCAAGCGCAGCCTCATCGGTGGCAAACTGTGACGCTGATTCAG 2820
2821 ATTTCAGCTTCAAGCGCTTCAAGCGCAGCCTCATCGGTGGCAAACTGTGACGCTTCA 2880
2821 ATTTCAGCTTCAAGCGCTTCAAGCGCAGCCTCATCGGTGGCAAACTGTGACGCTTCA 2880
2881 AATGCGCTGCAACCGTTTCAGACGCGATTTGGG 2914
2881 AATGCGCTGCAACCGTTTCAGACGCGATTTGGG 2914

RESULT 2

US-09-252-991A-8258
; Sequence 8258, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8258
; LENGTH: 3414
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8258

Query Match 4.3%; Score 125.2; DB 4; Length 3414;
Best Local Similarity 48.2%; Pred. No. 3.2e-24;
Matches 466; Conservative 0; Mismatches 473; Indels 27; Gaps 3;
QY 1307 CGGCGATTTGAAGCGCTTGAAGATATAATTCCTTTATTTTCAAGAGCTTGTGTTGACTTA 1366
Db 222 CGGCGATTTGCGCGGCTGATCGAGAGCTGATACATCGCCGACCTCGGCGTGAACAC 281
QY 1367 TCTGCACCTGATGCGCTGTTTAAATGCTGAAGGCAAAAGCAGCGGCGCTATGCGGT 1426
Db 282 TCTCTGGCTGCTGCGGT-----CTACCCGCTGCGCAGCGCGCTACGACAT 335
QY 1427 CAGCAGCTACCGCGATTCATTCGGGCTGCGGCAATAGCGACTTTCGCGGAAGTCAT 1486
Db 336 CGCCAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 395
QY 1487 TGCTGCGCTGCAAGCAGCGGCAATTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1546
Db 396 CGCGAGGCGCATCGACGCGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 455
QY 1547 CAACGACACGAAATGGCGCAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1606
Db 456 CGACGAGCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 515
QY 1607 CTATATTTTCCCGACCGCGGATGCGGCAATACGACGCACTTCGCGGCAATCTT 1666
Db 516 CTACTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 572
QY 1667 CCCGACGACGACCGCGGCGCTTCTGCAACTGGAAGCAGCGCTGCGGTGTCAGCAG 1726
Db 573 CGACACGAGCAGTCCAACTGCGACTTGGGACCGCGTACGCCCAAGTACTACTGCGACCG 632
QY 1727 CTTCATTTCTTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCGCGCAATGGC 1786
Db 633 CTTCATTTCTTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCGCGCAATGGC 692
QY 1787 GGGCGAAATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1846
Db 693 GGGGTGATGCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 752
QY 1847 CTTCATTTTGGAAACAAATGGGCAAGCTGCGAAACCTGCGGAGCGGCGACCGCTCAT 1906
Db 753 CTACCTGATCGAACCGGAGGCAACGAGGAGGAGAACTTGCAGGAGACCCACCGAGTCT 812
QY 1907 CGCGCGTTTCAATGCGCTTATGCGTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
Db 813 CAAGCGCATCCCGCGCGGAGCTGCGCACTATCCCGACCGCATGCTGCTGCGCGAGCG 872
QY 1967 CATGCTCCACCGCGACGAGTCTGCTCAATATATCGG-----GCAGCA 2008
Db 873 CAACGAGTGGCGGAGAGACACCGCGCGCTACTTCGCGCGGCGAGGATGGCGGAGGGCA 932
QY 2009 CGAATGCCAAATCGGTTACAAACCCCTGCAATGGCATTTGTTGGAAACACCCCTTGCAC 2068
Db 933 CGAATGCCAAATCGGTTACAAACCCCTGCAATGGCATTTGTTGGAAACACCCCTTGCAC 992
QY 2069 GCGCGAATCAACCTGCTCCATCAGCGCTGACCTACCGCGCACACCTGCGCGGAGCATAC 2128

Db 993 GGAGGATCGCTATCCGATCAGCAGCATCTCGGCCAGACCCCGGACATCCCGGCCAATTG 1052
Qy 2129 CGCCTGGTCACTAGTCCGAGCCAGCAGCATCGGCTGAGCTTTGCCGATGAAGA 2188
Db 1053 CCAATGGGGGATCTTCTCGCAACACACGAGCTGACCTCGAGATGGTCAACCGACGA 1112
Qy 2189 CGCGGCATATCTGGGATAGCGCTACGACACACCGCCCAATTCCTCAACCGGTCTTCT 2248
Db 1113 CGAGCGGACTATCTCTGGACCACTATGCGCGCGACCGCGCGCGCTCAACCTGGG 1172
Qy 2249 CAACCG 2254
Db 1173 CATCG 1178

RESULT 3
US-09-252-991A-8065/c
; Sequence 8065, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8065
; LENGTH: 4125
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8065

Query Match 4.2%; Score 125.2; DB 4; Length 4125;
Best Local Similarity 48.2%; Pred. No. 3.5e-24;
Matches 466; Conservative 0; Mismatches 473; Indels 27; Gaps 3;

Qy 1307 CGCGGATTTGAGGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTTGACTTA 1366
Db 3670 CGCGGACTTCGCGGGCTGATGAGAAGCTCGACTACATCGCGGACTCGGGGTGAACAC 3611
Qy 1367 TCTGACCTGATCGCGCTGTTTAAATGCTTGAAGCAAGGCGGGCTATGGGT 1426
Db 3610 TCTGCTGCTGCGGT-----CTACCGGTGCGGACCGCGGCGGCTACGACAT 3557
Qy 1427 CAGCAGCTACCGGATGTAATCGGCACTGGGCAATAGGCGACTTGGCGGAGTCAT 1486
Db 3556 CGCCAGTACGGTGGGTGACAGCGACTACGCGAGCTCGCGGCGCGGCTCAT 3497
Qy 1487 TGCTGGCTGACGAAGCGCGGATTTCCGCGGTGCTGATTTTATCTTCAACACACCTC 1546
Db 3496 CGCGAGGCGCATCGACCGGCTGCGGGTGATACCGAGCTGGTGATCAACACACCTC 3437
Qy 1547 CAGGACACGATGGGCGCAACGCTCGCGCGCGCGGCGCGCTTTTGGACAATTCTA 1606
Db 3436 CGACGACATCCCTGGTTTATCCCGCGCGCGCGCGGAGGATCGCGCGCGCGCG 3377
Qy 1607 CTATATTTTCCCGACCGCGGATGCGCGCAATACGACCGCACTCGCGGAAATCTT 1666
Db 3376 CTACTAGTCTGGTTCGACAGCG---ACGAGAAATACCGGCTACGCGGATCATCTTCAT 3320
Qy 1667 CCGGACGACACCGCGCGCTTCTGCACTGGAGAGCGGCTGGGTGGGACGAC 1726
Db 3319 CGACACGAGCGATCCCACTGGAGCTGGGACCGCGGTAGCCCAACAGTACTTGGACCG 3260
Qy 1727 CTTCAATCTTCAATGGGACTTGAATTAAGCAACCGGTGGGTATTCGCGCAATGGC 1786
Db 3259 CTTCTATTCACACCGCGGACCGCTGACTTGCACACCGCGAGGTCTCTGCGGAGGTGCT 3200

Qy 1787 GGGGAAATGCTTCTTTCCTTGCACACTTGGGCGTTTGACATCTCGGTATGSGATGGGTTGC 1846
Db 3139 CGGGTGTGCTGCTACTGCTGACATGGCGCTGCGCTGCGCTGCGCTGCGGATTC 3140
Qy 1847 CTTATTTGGAACAAATGGGGAAGAGTGGGAAACCTGCGGAGCGGCGGCGCTCAT 1906
Db 3139 CTACCTGTGAAACGCGGACCGGACCGGAGAACCTGCGGAGAGCCGACAGGTCT 3080
Qy 1907 CGCGGCTTCAATGCGGTTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAAAATCCGAAAGC 1965
Db 3079 CAAGCGGATCGCGCGGAGCTGGAGCGGCACTATCCCGACCGCATGCTGCTGGCGGAGC 3020
Qy 1967 CATGCTCCACCCCGGACCAAGTCTGTCGAATACATGG-----GCAGA 2008
Db 3019 CAACGAGTGGCGGAGAGACACCGCGCGCTACTTTCGGCGGCGAGGATGGCGGAGGGCGA 2960
Qy 2009 CGATGCCAAATCGGTTACAAACCCCTGCAATGCAATGTTGTGAAACACCTTGGCCAC 2068
Db 2959 CGATGCCCAATGCGCTTCCACTTCCGCTGATGCGCGGCACTGATGCGCATCGCCCA 2900
Qy 2069 GCGGAAATCAACCTGCTTCCATCAGGCGCTGACCTACCGGCGACACCTGCGCGAGCATAC 2128
Db 2899 GGAGGATCGTATCGATCAGACATCTGCGGCGAGACCCCGGACATCCCGGCGCAATTG 2840
Qy 2129 CGCCTGGTCACTAGTCCGCGCGGCGGACGACATCGGCTGGAGCTTTCGCGATGAAGA 2188
Db 2839 CCAATGGGCGATCTTCTTGGCGAACGACGAGCTGACCTGGAGATGGTACCGGCGA 2780
Qy 2189 CGCGCATATCTGGGCAATAAGCGGTACGACACCGCAATTCCTCAACCGGCTTCTTCT 2248
Db 2779 CGAGCGGACTATCTCTGGAACCACTATGCGCGGACCGCGCGCGCTCAACCTGGG 2720
Qy 2249 CAACCG 2254
Db 2719 CATCG 2714

RESULT 4
US-08-528-199-2
; Sequence 2, Application US/08528199
; Patent No. 5763228
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING
; TITLE OF INVENTION: MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/528,199
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,126
; FILING DATE: 14-SEP-1995
; APPLICATION NUMBER: US 08/485,126
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 156399/1994
; FILING DATE: 16-JUN-1994
; ATTORNEY/AGENT INFORMATION:

D5 485 GCAGTACTACTGGACCGCTTCTTCCACACAGCCGCGACCTGAACTTCGACACCCGA 544
QY 1768 GGGTATTCCCGCAATGGCGGGGAAATGCTGTCTTCCCTGCAATCTGGCGTTGACATCC 1827
D5 545 AGGTCCAGGACGCACTGCTGGAGCGATGCGTCTGCTGCGACATGGCGCTCGACGGCT 604
QY 1828 TGGTATGGATGGGTGCGCTTATTATTGGAAACAAATGGGGACAAGCTCGGAAACCTGC 1887
D5 605 TCGGCTCGACGGGTGCCCTACTCTACAGAGGTCGCCGCAACCAAGCGGAGAACCTCC 664
QY 1888 CGCAGGCGCACGCCCTCATCCGGCGGTTCATGCGCGTTATGCGTATTGCCGCGCCGCG 1947
D5 665 CCGAGAGCGACAGATGCTCAAGGGGTGCGGCGCTTCGTCGACGACAATACCCCGACC 724
QY 1948 TGTCTTCAAATCGAAGCATGCTCCACCCGACCAAGTCGTCCCAATACATCGGAGG 2007
D5 725 GGGTGTGCTGTACGAGGCGAACCAGTGGCCGACCGACGCTGGTGAGTACTTCGGGCGCG 784
QY 2008 ACGAATGCCAAATCGG 2023
D5 785 AGGAGCGTGAGGACGG 800

RESULT 6

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.4%; Score 99.4; DB 3; Length 4403765;
Best Local Similarity 49.2%; Pred. No. 9.9e-16;
Matches 293; Conservative 0; Mismatches 296; Indels 6; Gaps 1;
QY 1408 GCGACGGCGGTATGCGGTGACGAGTACCGGATGTCATTCGGCACTGGGCAACAATAG 1467
D5 152780 GCGACGGCGGTACGACATTCGGGACTTCTACAGTGTCTGCCGATTCGGCACCGTCG 152839
QY 1468 GCGACTTGGCGGAAGTCAATGCTGCGCTGCACGAAGCCGGCATTTCCGCGTGTGATTT 1527
D5 152840 ACGATTTGTCGCCCTGGTCGACGCGCTCACCGCGAGGTATCCGCATCATCACCAGCC 152899
QY 1528 TTATCTTCAACACACCTCCAAACGAAACGAAATGGGGCAACGCTGCGCGCGCGGAC 1587
D5 152900 TGGTATGATATCAACATCTGGAGTTCGACGCCCTGGTTTCAGAGTTCGCGGACCCAG 152959
QY 1588 OGCTTTTCGACAAFTTCTACTATATTTTCCCGACCGCGGATGCGCAATACGAC 1647
D5 152960 ACGACCGGTACGGTGAATATACGCTGTGGAGCGACACAGCGAGCGCTACACGACGCC 153019
QY 1648 GCACCTTCGCGGAATCTTCCCGACCGACCGCGGCGCTTCTCGCAACTGGAAGCG 1707
D5 153020 GGATCATCTTGG-----TCGACACCGAAGAGTGAATGCTGTCTTCCCAACTTGGCGCTTGCATCC 153073

QY 1708 GACGCTGGGTGTGAGGACCTTCAATTCCTTCATGGGACTTGAATTACAGCAACCCGT 1767
D5 153074 GACATTTCTACTGGCACCGAATTTCTTCTCCCAACCAAGATGAACTACGACCAACCCCG 153133
QY 1768 GGGTATTCCCGCAATGGCGGGGAAATGCTGTCTTCCCTGCAATCTGGCGTTGACATCC 1827
D5 153134 CCGTCAAGAGCGGATGATCGAGCTCAICGCGCTTTTGGCTGGCTTGGGCAATCGACGGT 153193
QY 1828 TCGTATGATGCGGTGCTTATTATTGGAAACAAATGGGGACAAGCTCGGAAACCTGC 1887
D5 153194 TTCGGTGGACCGGTGCCCTATCTCTTTGAAACGTGAGGGCCCAACTCGGAGAACCTGC 153253
QY 1888 CCGAGGCGCACGCCCTCATCCGGCGGTTCATGCGCTTATGCGTATTGCCGCGCCGCG 1947
D5 153254 CCGAAACACACGCTTTCTCAAGCGAGTCCGCAAGTGTGTGACGCAATTTCCCGGCG 153313
QY 1948 TGTCTTCAAATCCGAAGCATGCTCCACCCGACCAAGTCGTCCCAATACATCGG 2002
D5 153314 GGGTGTGCTAGCCGAAGCAATCAGTGGCGCGGCGATGTGCTCGAATATTTCGG 153368

RESULT 7

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 3.4%; Score 99.4; DB 3; Length 4411529;
Best Local Similarity 49.2%; Pred. No. 9.9e-16;
Matches 293; Conservative 0; Mismatches 296; Indels 6; Gaps 1;
QY 1408 GCGACGGCGCTATCGGTGACGAGTACCGGATGTCATTCGGCACTGGGCAACAATAG 1467
D5 152611 GCGACGGCGGTACGACATTCGGGACTTCTACAGTGTCTGCCGATTCGGCACCGTCG 152670
QY 1468 GCGACTTGGCGGAAGTCAATGCTGCGCTGCACGAAGCCGGCATTTCCGCGTGTGATTT 1527
D5 152671 ACGATTTGTCGCCCTGGTCGACGCGCTCACCGCGAGGTATCCGCATCATCACCAGCC 152730
QY 1528 TTATCTTCAACACACCTCCAAACGAAACGAAATGGGGCAACGCTGCGCGCGCGGAC 1587
D5 152731 TGGTATGATATCAACATCTGGAGTTCGACGCCCTGGTTTCAGAGTTCGCGGACCCAG 152790
QY 1588 CGCTTTTCGACAAFTTCTACTATATTTTCCCGACCGCGGATGCGCAATACGAC 1647
D5 152791 ACGACCGGTACGGTGAATATACGCTGTGGAGCGACACAGCGAGCGCTACACGACGCC 152850
QY 1648 GCACCTTCGCGGAATCTTCCCGACCGACCGCGGCGCTTCTCGCAACTGGAAGCG 1707
D5 152851 GGATCATCTTGG-----TCGACACCGAAGAGTGAATGCTGTCTTCCCGCTTGCCTCCG 152904
QY 1708 GACGCTGGGTGTGAGGACCTTCAATTCCTTCATGGGACTTGAATTACAGCAACCCGT 1767
D5 152905 GACAGTTCTACTGGCACCGAATTTCTTCTCCCAACACCGGATCTGAACTAGCAACCCCG 152964
QY 1768 GGGTATTCGCGCAATGCGCGCGGAAATGCTGTCTTCCCAACTTGGCGCTTGCATCC 1827

Db 152965 CCGTGAAGAGCGGATGATGACGTATCCCGCTTTGGCTGGGCAATCGACGGGT 153024
Qy 1828 TCGGTATGATGCGGTGCTTATTTGGAAACAAATGGGGACAAGCTGCGAAAACCTGC 1887
Db 153025 TCGGTGGAGCGGTGCGCTATCTTTGAAGCTGAGGGGACCACTGCGAGAACTGC 153084
Qy 1888 CGGAGGCGACGCGCTCATCGCGGCTTCAATGCGGTTATGCGGTATGCGCGCGCGCG 1947
Db 153085 CGGAACACACGCTTTTCTCAAGCGAGTCCGCAAGGTGGTGGAGCGAATTCGCCGCGCC 153144
Qy 1948 TGTCTTCAATCCGAAGCCATCGTCCACCGCGGACCAAGTGTCCATACATCGG 2002
Db 153145 GGGTGTCTAGCGAAGCCCAATGTCGGCGGGGATGTCTCGAATATTTCCG 153199

RESULT 8

US-09-107-532A-2141/c
; Sequence 2141, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2141:
SEQUENCE CHARACTERISTICS:
LENGTH: 768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...768
SEQUENCE DESCRIPTION: SEQ ID NO: 2141:
US-09-107-532A-2141
Query Match 3.0%; Score 86.6; DB 4; Length 768;
Best Local Similarity 45.8%; Pred. No. 5,9e-14;
Matches 299; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

Qy 11 TTCCGAACGAAACGTGATGCTTTGAGCCGAACACCTGTCCGGCAAGCGGCTACCGCCC 70
Db 691 TTAAGGAACCAACGTCATGACGGATCTGTCTGTTCCGCCAGCCAGACGAGAATTTTTC 632
Qy 71 CTTTTCGCCCATCGACATCGTAACATCGTTTGGTGGCAAGCTCTTTCGCTTTGAGCG 130
Db 631 CCAGATCGCCCATAGACATCGTAATGATAGGCATGGAGACAAAGCAATTTGCCGCTTAG 572
Qy 131 TGGCAGAAACAAAGTCAGCAAGTCTTTCGCGGCTTTTCGCGCATCACCGCAATTTTCAGA 190
Db 571 TCAACTCCATCAACCGCAAAACATCTCGTTCAATTTATTTGGAGTCACCTGCCAATTTCCGA 512
Qy 191 TGTCCGCGCGGAGTCTCTCATCTGTTTCAGACGGCATAGATTTCTTTCGGGCGCG 250
Db 511 TATCAGCGCCCAAAATGTTCCATGACATTCAGTCGGTAAAGCAAGACAGGCTCAGCTGGTG 452
Qy 251 TGGCGTGAAACTCATGATTTGCAGACGAGGCGCGGATGCCGTTTTTTTGGACATCGGCCA 310
Db 451 TTTCTTTAAATCATGACTACTGATAATTAAGGAATGTTTAAAGACTTTCATTTCTCTA 392
Qy 311 CGGCGCGCGGACGCGGTTTCGCCGGAATAAGCTCGATATCGATATGTCGGGAGGC 370
Db 391 TCAGTCCCTTTTCCTAGACATTTCTACTTTGAACAACTCGATATCGACCATATCTACTAATC 332
Qy 371 GGGTTTCAATCAGCGAGTCGAGCAGTTCAAAATAAATCGTCGGAACACGCGGAACGAGC 430
Db 331 CTGTGTATCATCATGTTTATAGAGTTCGGATATCGACGATCGAAATGTTTGTGTTCTC 272
Qy 431 CGCTTTCGCGCATGCCGTCTGAACGTAACAGCAGCGGCTTGTCCGGCAGCGCGTCGCGGA 490
Db 271 CGCCTTCAGCTAAGTACGAAACGTAATAATAACGCGCAGTCCATCAAGCATCTCATGGA 212
Qy 491 CGGCTCGGTGTGGCGCAATACCTTCGCGATGCTGCCCGGCATTCGCAAAATCGGCGC 550
Db 211 TAAGAACCAATGTTCTGCAACTTCATCAAGAAACCTTACGATAATGATCGACTC 152
Qy 551 GGAACCTCGACGATATCGAAGGCGAGGTTTTTGTGTTGGTCAAGTACGCGGGAAGTACGG 610
Db 151 GCCATTCATAAGTCGCGCCCTGCTTCAGTCGAAGACTGCTTCAAGTAATCTTCT 92
Qy 611 CGGCATCGCGGCGACAGCGGCAACGGGATTTTGGTGGCTCGCTTCCGATA 663
Db 91 CTCGAGCGGCTCCGATAAGAGCGCGCAAAATCTTTGGTCTACCGCTTCTCTATA 39

RESULT 9

US-09-328-352-3264/c
; Sequence 3264, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3264
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3264

Query Match 2.9%; Score 84.6; DB 4; Length 873;
Best Local Similarity 47.8%; Pred. No. 2.2e-13;
Matches 246; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

Qy 52 GCAAGCGGCTGACCGCCCTTTTCGCCCATCGATCGTACCAATCGGTTGGTGCA 111
Db 752 GCACACGAGAAATCGTACCTAATTTGCCCATCGACATTCATATATGCTTAGTGT 693
Qy 112 AGCTCTTTCGCTTTGAGCGGTGGCAGAAAGCAAAAGTCAGCAGCTCTTCGCGCTTTGCGGC 171

Db 692 TCGTGGCTTACTTTAAAGTTCGGTTTCATTAAAGTGAAGACATCTGTGTTAGATTGAGC 633
Qy 172 ATACCGCAATTTGAGATGTCGGCGCGAGTCTCTCACTGTTTCAGACGGCATAGC 231
Db 632 ATTACAGCAATTTTGAGAACATCTCGGCCCACTTTGATCTGTTTAAACAAGCGTTTTTCA 573
Qy 232 ATTCTCTCTGCGCGCGTGGTGAACACTCATGATTGACAGCAGCGCGCGATGCG 291
Db 572 ATTTCATCTTGGCTTGGTGTGTTTTTGGAGTCTGATTAGACATGACAAAGCACTTTC 513
Qy 292 TTTTGTGAGCATGCGCACCGCGCGCGAGCGGGTTTTCGCCGAAAGTTCGATA 351
Db 512 TTTTGTGATGCTGTTGTAACAATTTCTGAAACTACTTTTGTATCAGGAACATTTCAACA 453
Qy 352 TCGATAATGTCGGCAGCGCGCTTCAATCAGGAGTCGAGCAGTCAAAATAAATGCG 411
Db 452 TCTAACCGCTCATGAACGGATTTTAAATAGCCTGATAAGTCTTCCATTAATCAGCA 393
Qy 412 TCGAACACGGGAACGAGCGCGCTTTCGCATGCGCTGAACTGAAACAGCAGCGCGTTG 471
Db 392 TCACTAAATTTCAAGCTGACCCCTTGGTTTTTGTAGCAATAGTGCATCATTTGGCTTA 333
Qy 472 TCGGCGAGCGCTCGGAGCGTCTGCGTGTGGCGCAATCTTCGCCGATGTCGCCGCG 531
Db 332 TTGCTCAAAATCTTTTCAACTCATGCGCCCAAGGCAATCACTGTTTGTATGCTGGCA 273
Qy 532 CATTCCAAAATCGCGCGGAACTCGACGATATC 566
Db 272 AGCTTAATAATCGATACGGAACCTGCTAATC 238

RESULT 10

US-08-537-002A-4
; Sequence 4, Application US/08537002A
; Patent No. 5773282
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,002A
; FILING DATE: 29-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP No. 5773282 yet received
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2889 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-537-002A-4
Query Match 2.5%; Score 73.6; DB 1; Length 2889;
Best Local Similarity 47.0%; Pred No. 3.9e-10;
Matches 448; Conservative 0; Mismatches 479; Indels 27; Gaps 6;
Qy 1307 CGGCGATTTGAAGGGCTTGAAGATAAAATTCCTTATTTTCAAGAGTGTGTTGACTTA 1366
Db 81 GGGGACTTTTGAGGCGCTGAGCGGAAGCTTCCCTACTCTGGAGAGCTCGGGGTCAACAC 140
Qy 1367 TCTGCACCTGATGCGGTGTTTAAATGCCCTGAAGGCAAAAGCGACGCGCTATGCGGT 1426
Db 141 CCTCTGGCTCAAGCCCTTCTTCCAGTCCCC-----CTTGAGGACGACGGGTACGATAT 194
Qy 1427 CAGCAGCTTACCGGATGTCAATCCGGCACTGGGGCACAATAGCGGACTTGGCGAAGTCAT 1486
Db 195 CTCGACTTACTACAGATCTCTCCCGTCCACGGACCTCGAGGACTTCACCG-----T 248
Qy 1487 TGCTGCGTGCACAGAGCGGCAATTCGCGCTGCTGATTTTATCTTCAACACACATCT 1546
Db 249 GGRGAGGCGCCACGCGCGGGGATGAAGGTGATCATTTAGCTCTGCTGAAACCAACCTC 308
Qy 1547 CAACGACACGAATGGGCGCAACGCTGCGCGCGCGGACCGCTTTTCGACAAATTTCTA 1606
Db 309 CATTGACCACTTGGTTCCAG-----GAGCGGAGGAAGCGAATAGCCCCATCGCGGA 362
Qy 1607 CTATATTTTCCCGACCGCGGATGCGGACCAATAGACCGCACCTTCGCGGAATCTT 1666
Db 363 CT---GGTACCTGTGAGCGGACACCCCGGAGAGTACAAAGGGGTCCGGGTCTCTTCAA 419
Qy 1667 CCCGACCGACACCGCGCGCTTCTCGCACTGGAAGACGACGCTGGGTGTGACGAC 1726
Db 420 GGACTTTGAACCTTCAACTGGACCTTTGACCCCGTGGCCAAAGGCTTACTTGGACCG 479
Qy 1727 CTTCAATTTCTTCCAAATGGGACTTGAATTAAGCAACCCCGTGGGTATTTCCGCGCAATGGC 1786
Db 480 CTTCTACTGCGCACCGACCGACCTCACTGGGACAGCCCCGAGGTGGAGAAGGCGCATCCA 539
Qy 1787 GGGGAATGCTTTCCTTGGCAACTTGGGGTGTGACATCTGCGGTATGATGCGGTTC 1846
Db 540 CAGGTCAATGTTCTTGGCGCGACCTGGGGGTGACGCGCTTCCGCTTGGAGCGCATCC 599
Qy 1847 CTTTATTTGAAACAAATGGGACAGCTGCAAAACCTGCGGAGGCGC---AGCGCT 1903
Db 600 CTACCTCTACGAGCGGGAGGGGACCTCTCTGAGAGAACCTCCCGAGACCAATTGAGCGGT 659
Qy 1904 CATCGCGGCTTCAATGCGGTATGCGTATTTGCGCGCGCGCGCTGTGTTCTTCAATCCGA 1963
Db 660 GAAGCGCTGAGGAGGCGCTTGGAGGAGCGCTAGCGCCCCGGGAGAGATCTCTCTCGCGCA 719
Qy 1964 AGCATCTGTCACCCCGACCAAGTCTTCAATATATCGGGCA---GGACGAATGCCAAT 2020
Db 720 GGCCAAATGTGGCGGAGGAGACCTCTCTTCTGCGGACGGGGAGCGGGTCCACAT 779
Qy 2021 CGTTTAAACCCCTGCAAAATGGGATTTGTTGTGAACACCTTGGCCACCGCGGAGTCAA 2080
Db 780 GGCCTACAACTTCCCTCTGATGCCCGGATCTTATGSCCTTAAGGCGGGAGGACCGGG 839
Qy 2081 CTTGCTCCATCAGCGGCTGACCTACCGCACAACTGCGCGGAGCATACCGCTGGGTCAA 2140
Db 840 TCCCATTTGAACCATGCTCAGGAGGGGAGGGGATCCCGAAACCGCCAGTGGGCT 899
Qy 2141 CTACGTCCGAGCCACGACGATGCGGTGAGCGTTTCCGATGAGAGCGGGCATCT 2200
Db 900 CTTCTCCGCAACACGACGAGCTCACCTGAGAGAGGTACGAGGAGGAGCGGGAT 959
Qy 2201 GGGCATAGCGGCTACGACACCGCCCAATTCCTCAACCGCTTCTTCGTCAACCG 2254

Db 960 CATGTACAGGCTTACGCCCCGACCCCAAGTTCCGATCACTGGGGATCCG 1013

RESULT 11

US-08-863-010-4
; Sequence 4, Application US/08863010
; Patent No. 6067146
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,010
; FILING DATE:

CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,002
; FILING DATE: 29-SEP-1995
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP No. 6087146 yet received

FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI=1

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528

TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
; LENGTH: 2889 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: cDNA
; MISMATCHES: 0; Mismatches 479; Indels 27; Gaps 6;

Query Match 2.5%; Score 73.6; DB 3; Length 2889;
Best Local Similarity 47.0%; Pred. No. 3.9e-10;
Matches 448; Conservative 0;

US-08-863-010-4

QY 1307 CGCGCATTTGAAGGGCTTGAAGATATAATTCCTTTATTTCAAGAGCTTGGTTTGCATTA 1366
DB 81 CGGGGACTTTGAGGCTTGAGCGGAGCTTCCCTACCTGGAGGAGCTCGGGTCAACAC 140
QY 1367 TCTGCACCTGATGCGCTTTTAAATGCGCTGAAGGCAAAAGCGAGCGGGCTATGCGGT 1426
DB 141 CCTCTGGCTCATGCCCTTCTCCAGTCCCTCCCTGAGGAGCTGAGGAGTACGATAT 194
QY 1427 CAGCAGCTACCGGATGTCAATTCGGGCACTGGGCAAAATAGCGGACTTGGCGGAATCAT 1486
DB 195 CTCGCACTACTACGATCTCCCGTCCACGGGACCTGGAGGACTTCACCG-----T 248
QY 1487 TCGTGGCTGCACGAGCGGCAATTTCCGCCGCTGTCGATTTTATCTTCAACACACACCTC 1546

Db 249 GGACGAGGCCACGCGCGGGGATGAGGTGATCATTTGAGCTGCTGTAACACACCTC 308
QY 1547 CAACGAACAGAAATGGCGGCAACGCTGCGCGCGGGGACCCGCTTTTTCAGCAATTTCTA 1606
Db 309 CATTGACCACTCTGGTTCCAG-----GAGGCGAGGAAGCCGAATAGCCCCATGCGGGA 362
QY 1607 CTATATTTTCCCGACGCGCGGATGCCGACCAATACGACCGCACCTGCGCGGAAATCTTT 1666
Db 363 CT---GGTACGTGTGAGCGACACCCCGGAGAGTACAGGGGTCCGGGTCTCTTTCAA 419
QY 1667 CCCGACCAACGACCGCGGGGCTTTTCGCAACTGGAAGAGAGGAGCTGGGGTGGAGAC 1726
Db 420 GGACTTTGAAACCTTCAACTGGACCTTTGACCCCGCTGCGCAAGSCCTACTACTGGCAC 479
QY 1727 GTTCAATTCCTCCANTGGGACTTGATTAACAGCAACCCGCTGGGTATTCGCGCAATGGC 1786
Db 480 CTTCTACTGGACACGCGCGGCTCACTGGGACAGCCCGGAGTGGAGAGGCCATCCA 539
QY 1787 GGGCGAAATGCTGTTCCTTTCGCAACTTGGGGCTTGACATCTGCGTATGATGCGGTGC 1846
Db 540 CCAGGTCAATGTCTTCTGCGCGGACCTGCGGCTGAGCGGTTCGCGTGGACGCCATCCC 599
QY 1847 GTTTATTTGAAACAAATGGGCAAGCTGGAACAACTGCGGAGCGC---ACGCCCT 1903
Db 600 CTACTCTACGAGCGGAGGAGGACCTCTCGGAGACCTCCCGAGACCATTTGAGGCGGT 659
QY 1904 CATCCGGCGGCTTCAATGCCGTTATGCTATTGCGCGCGCCCGCTGTTCCTTCAAAATCGA 1963
Db 660 GAACGCTCTGAGGAAGGCCCTGGAGGAGCGCTACGCGCCCGGGAAGATCTCTCTCGCGA 719
QY 1964 AGCATGTGTCACCCCGACCACTGCTTCAATATACATCGGCA---GGACGAATGCCAAT 2020
Db 720 GGCACCAATGTGCGCGGAGGAGACCTCTCGGAGACCTCCCGGAGCGGGGTCCACAT 779
QY 2021 CGGTTACAAACCCCTGCAAAATGSCATTTGTGAAACACACCTTGCCTCCAGCGCGAAGTCAA 2080
Db 780 GGCCTACACTTCCCTCTGATGCGCGGATCTTCATGGCTTTCAGCGGAGGAGCGGGG 839
QY 2081 CCGTCTCATCAGGCGCTGACCTACCGGCAACACCTGCGGAGCATACCGCTCGGTCAA 2140
Db 840 TCCCATTTGAAACCATGCTCAAGGAGGCGGAGGATCCCGAAACCGCCAGTGGGCCCT 899
QY 2141 CTAGTCCGAGCGCACGACGACATCGGTGAGACCTTTCGATGAAGACGCGCATATCT 2200
Db 900 CTTCTCTCGCAACCAACGACGAGCTACCTGGAGAGTACGCGGAGGAGCGGGAGTT 959
QY 2201 GGGCATAGCGGCTACGACCAACCGCCCAATTCCTCAACCGCTTCTTCGTCAACCG 2254
Db 960 CATGTACGAGGCTTACGCCCCGACCCCAAGTTCCGATCAACCTGGGGATCCG 1013

RESULT 12

US-09-024-429-4
; Sequence 4, Application US/09024429
; Patent No. 6165768

GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/024,429
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/537,002
/ FILING DATE: 29-SEP-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 260984/1994
/ FILING DATE: 01-OCT-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 255829/1995
/ FILING DATE: 08-SEP-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: YUN, Allen C.
/ REGISTRATION NUMBER: 37,971
/ REFERENCE/DOCKET NUMBER: TSUSAKI=1B
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-628-5197
/ TELEFAX: 202-737-3528
/ TELEX: 248633
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2889 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ PS-09-024-429-4

```

Query Match	2.5%;	Score 73.6;	DB 3;	Length 2889;
Best Local Similarity	47.9%;	Pred. No. 3.9e-10;		
Matches 448;	Conservative	0;	Mismatches 479;	Indels 27; Gaps 6;
QY	1307	CGGCGATTTCGAAGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGTTGACTTA	1366	
DB	81	CGGGGACTTTGAGGCTTGAGCGGAGCTTCCTACCTGAGGAGCTCGGGGTCAACAC	140	
QY	1367	TCGCACTCATGCGCGCTGTTTAAATGCCCTTGAGGCAAAAGCAGCGCGCTATCGGT	1426	
DB	141	CCTCTGGCTCATGCGCTTCCTTCAGTCCCC-----CTTGAGGAGCAGCGGTACGATAT	194	
QY	1427	CAGCAGCTACCGGATGTCAATCGGCACATAGGGACATTTGGCGGAAGTCAT	1486	
DB	195	CTCGACTACTACAGATCTCTCCCGTCCAGCGACCTCTGAGGACTTCACCG-----T	248	
QY	1487	TGCTGCCTGCAGAAAGCGGCAATTCGCGCTGTCGATTTTATCTTAAACACACCTC	1546	
DB	249	GGACAGGCGCACGCGCGGGGATGAAGTGATCAATTGAGCTCGTCTGAACACACACCTC	308	
QY	1547	CAACGACAGAAATGGGCGCAACGCTGCGCGCGCGGACCGCGTTTCACACAAATTTCTA	1606	
DB	309	CATTGACCACTCTTGTTTCAG-----GAGGCGAGGAAGCGGATAGCCCCATGCGGA	362	
QY	1607	CTATATTTTCCCGACCCCGGATGCGCGACCAATACGACGCAACCTCTGCGGHAATCTT	1666	
DB	363	CT---GGTACGTGTGAGCGACACCCCGGAGAGTACAAAGGGGTCTCGGGTCATCTTCAA	419	
QY	1667	CCCCGACGACCCCGGGCGCTCTTCGCACTGGAAGAGCAAGCGCTGGGTGTGGAGCAC	1726	
DB	420	GGACTTTGAAACCTCCAACTGGACCTTTGACCCCGTGGCCAGGCTTACTTGGCACCG	479	
QY	1727	CTTCAATTCCTTCCAAATGGGACTTGAATATACGCAACCCGTGGGTAATTCGCGCGAATGCG	1786	
DB	480	CTTCTACTGGCACCCAGCCCGACCTCAACTGGGACAGGCCCGAGGTGGAGAGGCCATCCA	539	
QY	1787	GGGCGAATTCGTGTTCTTGGCCACTTGGCGTTTGACATCTGCTGATGATGCGGTTCG	1846	
DB	540	CCAGGTTCATGTTCTTCTGGGCGGACCTTGGGGGTGGAAGGCTTTCGCTTGAAGCGCATCCC	599	
QY	1847	CTTTATTTTGAAACAAATGGGGAACAAGCTGCGGAAACCTGCCGAGGCGC---ACGCCCT	1903	

```

1  RESULT 13
2  US-08-537-002A-5
3  Sequence 5, Application US/08537002A
4  Patent No. 5773282
5  GENERAL INFORMATION:
6  APPLICANT: TSUSAKI, Keiji
7  APPLICANT: KUBOTA, Michio
8  APPLICANT: SUGIMOTO, Toshiyuki
9  TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
10 TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
11 NUMBER OF SEQUENCES: 17
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: BROWDY AND NEIMARK
14 STREET: 419 Seventh Street, N.W., Suite 300
15 CITY: Washington
16 STATE: D.C.
17 COUNTRY: USA
18 ZIP: 20004
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Floppy disk
21 COMPUTER: IBM PC compatible
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23 SOFTWARE: PatentIn Release #1.0, Version #1.30
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/537,002A
26 FILING DATE: 29-SEP-1995
27 CLASSIFICATION: 435
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: JP 260984/1994
30 FILING DATE: 01-OCT-1994
31 PRIOR APPLICATION DATA:
32 APPLICATION NUMBER: JP No. 5773282 yet received
33 FILING DATE: 08-SEP-1995
34 ATTORNEY/AGENT INFORMATION:
35 NAME: BROWDY, Roger L.
36 REGISTRATION NUMBER: 25,618
37 REFERENCE/DOCKET NUMBER: TSUSAKI=1
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: 202-628-5197
40 TELEFAX: 202-737-3528
41 TELEX: 248633
42 INFORMATION FOR SEQ ID NO: 5:
43 SEQUENCE CHARACTERISTICS:
44 LENGTH: 3600 base pairs
45 TYPE: nucleic acid
46 STRANDEDNESS: DOUBLE

```

QY	2081	CTGTGCTCCATCAGGCGCTCAGCTACCGCCACAACCTGCGCCGAGCATACCGCCCTGGGCTCAA	2144
Db	1380	TCCCATGTGAACCATCTCTCAGGAGCGGAGGGGATCCCGGAAACCGCCCGAGTGGGCCCT	1439
QY	2141	CTAAGTCCGAGCAGCAGCAGCATCGCTGCGAGCTTTGCGCGATGAAGAGCGCGCATATCT	2200
Db	1440	CTTCTCTCGCGAACCAGCAGCAGGCTCACCTCGGAGAGAGTTCACGAGGAGGAGCGGGAGTT	1499
QY	2201	GGGCATAAGCGGCTACGAGCACCGCCCAATTCTCTCAACCGCTCTTCGCTCAACCG	2254
Db	1500	CATGTACGAGGCGCTACGCCCGCCGACCCCAAGTTCGCGATCAACCTGGGATCCG	1553

RESULT 14

US-08-863-010-5

; Sequence 5, Application US/08863010

; Patent No. 6087146

GENERAL INFORMATION:

APPLICANT: TSUSAKI, Keiji

APPLICANT: KUBOTA, Michio

APPLICANT: SUGIMOTO, Toshiyuki

TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR

TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/863,010

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/537,002

FILING DATE: 29-SEP-1995

APPLICATION NUMBER: JP 260984/1994

FILING DATE: 01-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP No. 6087146 yet received

FILING DATE: 08-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI=1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3600 base pairs

TYPE: nucleic acid

STRANDEDNESS: DOUBLE

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: *Thermus aquaticus*

INDIVIDUAL ISOLATE: ATCC 33923

FEATURE:

NAME/KEY: 5'UTR

LOCATION: 1..540

IDENTIFICATION METHOD: E

NAME/KEY: mat peptide

LOCATION: 541..3429

IDENTIFICATION METHOD: S

NAME/KEY: 3'UTR
LOCATION: 3430..3600
IDENTIFICATION METHOD: E
US-08-843-010-5

Query Match 2.5%; Score 73.6; DB 3; Length 3600;
Best Local Similarity 47.0%; Pred. No. 4.3e-10;
Matches 448; Conservative 0; Mismatches 479; Indels 27; Gaps 6;

QY 1307 CGGGGATTGAGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGTTGACTTA 1366
DB 621 CGGGGACTTTGAGGCTTGAAGAGCTTCCCTTCTGAGGAGCTCGGGGTCAACAC 680

QY 1367 TCTGACCTGATGCCGCTGTTTAAATGCTTGAAGGCAAAAGCGACGCGGCTATGCGGT 1426
DB 681 OCTGCGCTGATGCCCTTCTCCAGTCCCTCTGAGGAGCAAGCGGTACGATAT 734

QY 1427 CAGCAGCTACCGCATGTCAATCGGCACTGGCACAATGCGGACTTGGCGGAGTCAT 1486
DB 735 CTCGCACTACTACCAAGATCTCCCGCTCCAGGACCTGAGGAGCTTCCAG-----T 788

QY 1487 TGCTGCGCTGACGAGCGGCAATTCGCGCTGCTGCAATTTTATTTTCAACCAACATC 1546
DB 789 GGACGAGCGCCACGCGCGGGGATGAAGTGATCATTCAGCTCTGCTGAAACCAACATC 848

QY 1547 CAACGACACGAAATGGGCGCAACGCTGCGCGCGCGGACCGCTTTTCGACATTTCTA 1606
DB 849 CATTGACCACTTGGTTTCCAG-----GAGCGGAGGAGCGCAATAGCCCATGCGGA 902

QY 1607 CTATATTTTCCCGACCGCGGATGCCGACCAATACGACCGCACCTTCGCGCAATCTT 1666
DB 903 CT---GGTACGTGTGGAGGACACCCCGGAGAGTACAGGGGTCCGGTCACTTCAA 959

QY 1667 CCCGACGACGACCCGCGGGGCTTCTGCACTGGAAGACGAGCGCTGGGTGTGACAC 1726
DB 960 GGACTTTGAAACCTCAACTGACCTTTGACCGCTGCGCAAGGCTTACTTGGCACCG 1019

QY 1727 CTTCAATTCCTTCCAAATGGAGCTTCAATACGACCAACCGCTGGGTATTCCGCGCAATGGC 1786
DB 1020 CTTCTACTGGACGACCGGACTCTCAATGAGACAGCCCCGAGGCTGGAGAGGCAATCA 1079

QY 1787 GGGCAAAATGCTTCTTCCAAATGGCGCTTGCATCTGCAATCTGCTATGATGGGTGTC 1846
DB 1080 CCAGGTGATGTTCTTCTGGCGCGAGCTGGGGGTGACGCTTCCGCTGGAGCGCATCC 1139

QY 1847 CTTTATTTGGAACAAATGGGAGCAAGCTGCAAAACCTGCGGAGGCGC---ACGCCCT 1903
DB 1140 CTACCTCTACGAGCGGAGGAGGAGCTCTTCCGAGAACCTTCCCGAGACCATGAGCGGT 1199

QY 1904 CATCGCGGCTTCAATGCGTTATGCGTATTCGCGCGCGCGCTGTTCTTCAATCCGA 1963
DB 1200 GAAGCGCTGAGGAGGCGCTGAGGAGCGCTACGCGCGCGGAGAGATCTCTCGCGCA 1259

QY 1964 AGCCATGCTCACCGGACCAAGTCTGCAATACATCGGCA---GGACGATGCCAAT 2020
DB 1260 GGCACACATGTGCGCGGAGGAGACCTTCCCTTCTGCGGAGCGGAGCGGCTCCCAT 1319

QY 2021 CGGTTACAAACCCCTGCAAAATGCAATGTTGTGGAACACCTTTCGACGCGGAGTCAA 2080
DB 1320 GGCCTACAACTTCCCTCTGATGCCCGGATCTTCATGGCCCTAAGCGGAGGAGCGGG 1379

QY 2081 CTGCTTCCATAGGCGCTGACCTACCGCACAACTGCGCGGAGCATACGCGCTGGGTCA 2140
DB 1380 TCCCATTTGAACCTTCTCAAGGAGCGGAGGAGTCCCGAAACCGCGGAGTGGCGCT 1439

QY 2141 CTACGTCGCGACGACGACGATCGGCTGGAGCTTTGCGGATGAAGACGCGGATATCT 2200
DB 1440 CTTCTTCCGACCAACGACGAGCTCACCTTGGAGAGGTACGAGGAGGAGCGGAGTT 1499

QY 2201 GGGCATAAGCGCTAGACCAACCGCAATTCCTCAACCGCTTCTTGGTCAACCG 2254
DB 1500 CATGTACGAGGCTTACGCGCGGAGGAGGAGTTCGCTATCAACTCTGGGATCCG 1553

RESULT 15
US-09-024-429-5
Sequence 5, Application US/09024429
Patent No. 6155768
GENERAL INFORMATION:
APPLICANT: TSUSAKI, Keiji
APPLICANT: KUBOTA, Michio
APPLICANT: SUGIMOTO, Toshiyuki
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
TITLE OF INVENTION: CONVERTING MALTOSE INTO TRHALOSE
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/024,429
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/537,002
FILING DATE: 29-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 260984/1994
FILING DATE: 01-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 255829/1995
FILING DATE: 08-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: TSUSAKI=1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3600 base pairs
TYPE: nucleic acid
STRANDEDNESS: DOUBLE
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Thermus aquaticus
INDIVIDUAL ISOLATE: ATCC 33923
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..540
IDENTIFICATION METHOD: E
NAME/KEY: mat peptide
LOCATION: 541..3429
IDENTIFICATION METHOD: S
NAME/KEY: 3'UTR
LOCATION: 3430..3600
IDENTIFICATION METHOD: E
US-09-024-429-5

Query Match 2.5%; Score 73.6; DB 3; Length 3600;
Best Local Similarity 47.0%; Pred. No. 4.3e-10;
Matches 448; Conservative 0; Mismatches 479; Indels 27; Gaps 6;

QY 1307 CGGGGATTGAGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGTTGACTTA 1366
DB 621 CGGGGACTTTGAGGCTTGAAGAGCTTCCCTTCTGAGGAGCTCGGGGTCAACAC 680

QY 1367 TGTGACCTGATGCGCTGTTAAATGCGCTGAGGCAAAAGCGACGGCGCTATGCGGT 1426
 Db |||||
 681 CCTTGGCTCATGCTCTTCCAGTCCC-----CTTGAGGAGACGCGGTACGATAT 734
 QY 1427 CAGCAGCTACCGGAGTCAATCGGCACTGGGCAATAGCGGACTTGGCGGAAGTCA 1486
 Db |||||
 735 CTCGCACTACTACAGATCTCCCGCTCCACGGGACCTGGAGGACTTCAACG-----T 788
 QY 1487 TGCTGCGTGCACGAGCGGCAATTCGCGCTGCTGATTTTATCTTCAACCAACACCTC 1546
 Db |||||
 789 GGAGGAGGCGCCAGCGCGGGGATGAGGTGATCATTTGAGCTCGTCTGNACCAACCTC 848
 QY 1547 CAACGAACAGAAATGGGCGCAACGCTGCGCGCGGCGACCGCGCTTTTGGACAAATTTCTA 1606
 Db |||||
 849 CATTGACCACTCTTGGTTCCAG-----GAGCGAGGAAGCGCAATAGCCCATCGGGA 902
 QY 1607 CTATATTTTCCCGACCGCGGATGCGCGACCAATAGACCGCACCGCTGCGGGAATCTT 1666
 Db |||||
 903 CT---GGTACGTGTGAGCGACACCGCGGAGAGTACAAGGGGTCCGGGTCTATTCAA 959
 QY 1667 CCGGACCAAGCACCGCGGCTTCTCGCAACTGGAAGACGACGCTGGGTGTGGAACGAC 1726
 Db |||||
 960 GGACTTTGAACCTTCAACTGGACCTTTGACCCCGTGGCCCAAGGCCCTACTACTGGCAACG 1019
 QY 1727 CTTCAATTCCTTCCAAATGGGACTTGAATTAAGCAACCGTGGGTATTCGGGCGAATGSC 1786
 Db |||||
 1020 CTTCTACTGCAACCAACCGCGGACCTCAACTGGGACGCGCGGAGTGGAGAGGCCATCCA 1079
 QY 1787 GGGCGAAATGCTGCTTCTTCCAACTTGGGCTGTGACATCTCGGCTATGATGCGGTTCG 1846
 Db |||||
 1080 CAGGTCAATGTTCTTGTGGCGGACCTTGGGGTGGAGCGCTTCCGCTTGAACGCAATCCC 1139
 QY 1847 CTTTATTGGAAACAATGGGCAAGCTGGGAAACCTTGGCGGAGCGC---AGCCCT 1903
 Db |||||
 1140 CTACCTCTACGAGCGGAGGAGACCTCTCTCGAGAACCTTCCCGAGACCAATGAGCGGT 1199
 QY 1904 CATCGCGGCTTCAATGCGCTTATGGTATTGCGCGCGCGCGCTCTTCTTCAAAATCCGA 1963
 Db |||||
 1200 GAAGCGCTGAGGAGGCTTGGAGAGCGCTACGSCCCCGGGAAGATCTCTCTCGCGGA 1259
 QY 1964 AGCATGCTCCACCCCGACCAAGTCTTCCAAATACATCGGCA---GGAAGAAATGCCAAT 2020
 Db |||||
 1260 GGCCAAACATGTGGCGGAGGAGACCTCCCTACTTTCGGGACGCGGAGCGGGTCCACAT 1319
 QY 2021 CGGTTACAAACCCCTGCAATGSCATTTGTGGAACACCTTGGCAACCGCTTGGCAACGCGCAATCAA 2080
 Db |||||
 1320 GGCTTACAACTTCCCTCTGATGCCCGGATCTTCAATGGCCCTAAGCGGAGAGACCGGGG 1379
 QY 2081 CTTGCTCCATCAGCGCTGACCTACCGCAACACCTGCGCGAGCATACCGCTTGGGTCAA 2140
 Db |||||
 1380 TCCCATTTGAACCATGCTCAGGAGGCGGAGGGATCCCGGAACCGCCAGTGGGCGCT 1439
 QY 2141 CTACGTCGCGAGCAACGACATCGGCTGGAAGTTTGGCGATGAAGACGCGCATATCT 2200
 Db |||||
 1440 CTTCTCCGCAACCAACGACGAGCTCACCTTGGAGAGGTCAAGGAGGAGGAGGAGGAGTT 1499
 QY 2201 GGGCATAGCGGCTACGACCAACCGCAATTCCTCAACCGCTTCTTGTCAACCG 2254
 Db |||||
 1500 CATGTAGAGGCGCTACGCGCCCGGACCCCAAGTTCCGATCAACTGCGGATCCG 1553

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 16:01:07 ; Search time 886 Seconds
(without alignments)
10494.866 Million cell updates/sec

Title: US-09-843-007A-1

Perfect score: 2914
Sequencing: 1 GAGTTTCGTTCCGACACC.....CGTTTCAGCGCATTTGCG 2914

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2141354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubna/PTC_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubna/PTCUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubna/US09D_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2914	100.0	2914	9	US-09-843-007-1
2	169	5.8	759	9	US-09-815-242-6076
3	161.4	5.5	882	9	US-09-815-242-9839
4	142.6	4.9	801	9	US-09-815-242-6759
5	142.6	4.9	6415	10	US-09-070-927A-275
6	120.4	4.1	3303	9	US-09-815-242-7764
7	119	4.1	1716	14	US-10-156-761-2790
8	119	4.1	9025608	14	US-10-156-761-1
9	118	4.0	483	10	US-09-974-300-399
10	111.8	3.8	1794	10	US-09-738-626-2527
11	111.8	3.8	3309400	10	US-09-738-626-1
12	111	3.8	1749	14	US-10-156-761-7374
13	70.8	2.4	1599	14	US-10-156-761-1713
14	70.8	2.4	9025608	14	US-10-156-761-1
15	65	2.2	1589	10	US-09-974-300-657
16	59.8	2.1	1782	14	US-10-061-269-13

17	56.8	1.9	1704	14	US-10-156-761-6284
18	52	1.8	1683	10	US-09-974-300-655
19	47.2	1.6	520	14	US-10-184-644-332
20	47.2	1.6	520	14	US-10-184-644-332
21	47.2	1.6	1848	14	US-10-156-761-5957
22	46.8	1.6	1704	14	US-10-061-269-15
23	45.2	1.6	1703	10	US-09-974-300-644
24	44.2	1.5	1659	12	US-10-223-277-7
25	44.2	1.5	1659	12	US-10-223-277-7
26	43.4	1.5	1642	13	US-10-014-436-1
27	42.8	1.5	1344	14	US-10-156-761-6042
28	42.6	1.5	1653	9	US-09-815-242-9889
29	42.6	1.5	30246	8	US-08-781-986A-56
30	42.4	1.5	1521	14	US-10-156-761-6290
31	42.4	1.5	77536	12	US-09-940-316B-1
32	41.8	1.4	5418	14	US-10-156-761-5959
33	41.6	1.4	933	14	US-10-156-761-2119
34	41.6	1.4	933	14	US-10-156-761-2300
35	41.6	1.4	963	14	US-10-156-761-686
36	41.4	1.4	1872	12	US-10-084-843-17
37	41.4	1.4	1872	12	US-10-193-002-17
38	41.4	1.4	1872	12	US-10-098-732A-1
39	41.4	1.4	2151	14	US-10-127-032-72
40	41.4	1.4	4935	14	US-10-156-761-5059
41	41.2	1.4	2586	10	US-09-738-626-456
42	40.6	1.4	1785	10	US-09-738-626-2326
43	40.6	1.4	2325	14	US-10-156-761-2691
44	40.6	1.4	3033	14	US-10-212-219-5
45	40.2	1.4	1389	14	US-10-156-761-5431

ALIGNMENTS

RESULT 1

US-09-843-007-1

; Sequence 1, Application US/09843007

; Patent No. US20020092040A1

; GENERAL INFORMATION:

; APPLICANT: Rosemann, Jens

; Buttcher, Volker

; Weish, Thomas

; TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE

; OF FACILITATING THE SYNTHESIS OF LINEAR

; ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND

; MICROORGANISMS

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/843,007

; FILING DATE: 26-Apr-2001

; CLASSIFICATION: <Unknown>

; ERROR APPLICATION DATA:

; APPLICATION NUMBER: 08/737,752

; FILING DATE: <Unknown>

; APPLICATION NUMBER: DE P 44 47 388.5

; FILING DATE: 22-DEC-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: James F. Haley, Jr., Esq.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: GPB-1

; TELECOMMUNICATION INFORMATION:

Sequence 6284, Ap
Sequence 655, App
Sequence 332, App
Sequence 332, App
Sequence 5957, Ap
Sequence 15, Appl
Sequence 644, App
Sequence 7, Appl
Sequence 8, Appl
Sequence 1, Appl
Sequence 6042, Ap
Sequence 9889, Ap
Sequence 56, Appl
Sequence 6290, Ap
Sequence 1, Appl
Sequence 5959, Ap
Sequence 2119, Ap
Sequence 2300, Ap
Sequence 686, App
Sequence 17, Appl
Sequence 17, Appl
Sequence 1, Appl
Sequence 72, Appl
Sequence 5059, Ap
Sequence 456, App
Sequence 2326, Ap
Sequence 2891, Ap
Sequence 5, Appl
Sequence 5431, Ap

TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2914 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Neisseria polysacchara

IMMEDIATE SOURCE:

LIBRARY: genomic library in pBluescriptII SK

CLONE: pNB2

FEATURE:

NAME/KEY: CDS

LOCATION: 957..2867

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-843-007-1

Query Match 100.0%; Score 2914; DB 9; Length 2914;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2914; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAGTTTGGCTTCCGAAACGACGATGATGCTTGAGCCGACACACCTGTCGCGCAAGGGCG 60
DB 1 GAGTTTGGCTTCCGAAACGACGATGATGCTTGAGCCGACACACCTGTCGCGCAAGGGCG 60
QY 61 CTGACGCGCCCTTTTCCGCCATCGACATCGTAACAAATCGTTTGGTGGCAAGCTCTTTC 120
DB 61 CTGACGCGCCCTTTTCCGCCATCGACATCGTAACAAATCGTTTGGTGGCAAGCTCTTTC 120
QY 121 GCTTTGAGGTGGCAGAAACGAAAGTCAGCACTCTTCCGCGCTTTTCCGCGATCACCGCA 180
DB 121 GCTTTGAGGTGGCAGAAACGAAAGTCAGCACTCTTCCGCGCTTTTCCGCGATCACCGCA 180
QY 181 AHTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGAGCGCATACGATTTCTTCT 240
DB 181 AHTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGAGCGCATACGATTTCTTCT 240
QY 241 TCGCGCGGCTGGGGTGAACCTCATGATTCAGAGCAGGCGCGCGATGCGGTTTTTGA 300
DB 241 TCGCGCGGCTGGGGTGAACCTCATGATTCAGAGCAGGCGCGCGATGCGGTTTTTGA 300
QY 301 GCATGCGCACGCGCGCGCGAGCGGTTTCGCGGAAAGCTCGCATATCGATAATG 360
DB 301 GCATGCGCACGCGCGCGCGAGCGGTTTCGCGGAAAGCTCGCATATCGATAATG 360
QY 361 TCGCGCAGCGGCTTTCAATCAGCGATCGAGCTTCAAAATAATATGTCGCAACAC 420
DB 361 TCGCGCAGCGGCTTTCAATCAGCGATCGAGCTTCAAAATAATATGTCGCAACAC 420
QY 421 GGAACGAGCGGCTTGGCCATCGCTGAAACGTAACGAGCGGCTTTCGCGCAGC 480
DB 421 GGAACGAGCGGCTTGGCCATCGCTGAAACGTAACGAGCGGCTTTCGCGCAGC 480
QY 481 GCGTCGCGGAGCGGCTGCGGTGTCGCAATCTTCCGATGCTGCCCGCATTCGCAAA 540
DB 481 GCGTCGCGGAGCGGCTGCGGTGTCGCAATCTTCCGATGCTGCCCGCATTCGCAAA 540
QY 541 AAATCGCGCGGACTCGAGATATCGAGCGGAGGTTTTTGTATTTGGTCAAGTACGGG 600
DB 541 AAATCGCGCGGACTCGAGATATCGAGCGGAGGTTTTTGTATTTGGTCAAGTACGGG 600
QY 601 GAAAGTACGCGGCGCATCGCGGCGGCAAGAGCGGCAATTTTGGTGGCTTCGCTTCG 660
DB 601 GAAAGTACGCGGCGCATCGCGGCGGCAAGAGCGGCAATTTTGGTGGCTTCGCTTCG 660
QY 661 ATAAAGGTGTTTTGACGCTCAGCTGCTGATGCGGCTTGTTCGCGTGAAGAAC 720
DB 661 ATAAAGGTGTTTTGACGCTCAGCTGCTGATGCGGCTTGTTCGCGTGAAGAAC 720
```

```
QY 721 GGTAAAGACGCAATTATAGCAAGGCAAGCGCATGTTTCAGACGCGCATTTCTGTGGCG 780
DB 721 GGTAAAGACGCAATTATAGCAAGGCAAGCGCATGTTTCAGACGCGCATTTCTGTGGCG 780
QY 781 CGGCTTGATATGAATCAAGCAGCATCCGCATATCGGAATCGAGACTTTGGCACAAGCCCTG 840
DB 781 CGGCTTGATATGAATCAAGCAGCATCCGCATATCGGAATCGAGACTTTGGCACAAGCCCTG 840
QY 841 TCTTTTCTAGTCAGTCGCGGATTTCTTTCAGTATGATTCAGCAGCAGCCCTTACAGCGAT 900
DB 841 TCTTTTCTAGTCAGTCGCGGATTTCTTTCAGTATGATTCAGCAGCAGCCCTTACAGCGAT 900
QY 901 TTGCAAGATACGCGCAGACCGCGCTCGGAACTTCAGAAATCGGAGCAGCGATCATGT 960
DB 901 TTGCAAGATACGCGCAGACCGCGCTCGGAACTTCAGAAATCGGAGCAGCGATCATGT 960
QY 961 TGACCCCGACGCAAGTCTGGTTTGTATTTTACAGTACTTCAAAACAGCATCTTGACA 1020
DB 961 TGACCCCGACGCAAGTCTGGTTTGTATTTTACAGTACTTCAAAACAGCATCTTGACA 1020
QY 1021 TCTACAGCCCGAAACAGCGCGCGCATCGAAAAATCGAAAGACTGGCGCGAGTTTTCG 1080
DB 1021 TCTACAGCCCGAAACAGCGCGCGCATCGAAAAATCGAAAGACTGGCGCGAGTTTTCG 1080
QY 1081 GCGCATGATACGCAATTTCCCAAACTGATGAAAGCACTCGACAGCGGTGTACGCAACA 1140
DB 1081 GCGCATGATACGCAATTTCCCAAACTGATGAAAGCACTCGACAGCGGTGTACGCAACA 1140
QY 1141 ACGAGCCCTGCTGCTATGCTGGAAATCTGCTGGCGCAGCGCATGGCAAGCTATTCC 1200
DB 1141 ACGAGCCCTGCTGCTATGCTGGAAATCTGCTGGCGCAGCGCATGGCAAGCTATTCC 1200
QY 1201 AACGCAACTCATCTCTTAAAGATATCGATATCGGCGCGGCAAAACAAACCGCATTTG 1260
DB 1201 AACGCAACTCATCTCTTAAAGATATCGATATCGGCGCGGCAAAACAAACCGCATTTG 1260
QY 1261 TGTCCAACAAACAGTCGCGCGGCTGTCTAGTGTGTTGTTGCTGCGCGCATTTGAAG 1320
DB 1261 TGTCCAACAAACAGTCGCGCGGCTGTCTAGTGTGTTGTTGCTGCGCGCATTTGAAG 1320
QY 1321 GCTTGAAGATATAAATTCCTTTTCAAGAGCTTGGTTGACTTATCTGACCTGATGC 1380
DB 1321 GCTTGAAGATATAAATTCCTTTTCAAGAGCTTGGTTGACTTATCTGACCTGATGC 1380
QY 1381 CGCTGTTTAAATGCTCGAAGCAAAAGCGAGCGGCTATCGGTGAGCAGCTACCGG 1440
DB 1381 CGCTGTTTAAATGCTCGAAGCAAAAGCGAGCGGCTATCGGTGAGCAGCTACCGG 1440
QY 1441 ATGTCAAATCCGCACTGGGCAATATAGCGGACTTTCGCGGAAAGTCAATGCTGCGTGCAG 1500
DB 1441 ATGTCAAATCCGCACTGGGCAATATAGCGGACTTTCGCGGAAAGTCAATGCTGCGTGCAG 1500
QY 1501 AAGCGCGATTTCCGCGCTGCTGATTTTCTTCAACCAACCTTCTGCTGCTGCGG 1560
DB 1501 AAGCGCGATTTCCGCGCTGCTGATTTTCTTCAACCAACCTTCTGCTGCTGCGG 1560
QY 1561 GGGCGCAAGCTGCGCGCGCGGCTTTTCGACAAATTTCTACTATATTTTCCCG 1620
DB 1561 GGGCGCAAGCTGCGCGCGCGGCTTTTCGACAAATTTCTACTATATTTTCCCG 1620
QY 1621 ACCGCGGATGCCGCAACCAATAAGCAGCCGCACTTCGCGGAAATCTTCCCGGACGAGC 1680
DB 1621 ACCGCGGATGCCGCAACCAATAAGCAGCCGCACTTCGCGGAAATCTTCCCGGACGAGC 1680
QY 1681 CGGGCGGCTTCTCGCAACTGGAAGCAGCGCTGCTGGAGCTTCAATTTCTTCC 1740
DB 1681 CGGGCGGCTTCTCGCAACTGGAAGCAGCGCTGCTGGAGCTTCAATTTCTTCC 1740
QY 1741 AATGGAGCTTGAATTCAGCAACCGCGTGGTATTCGCGCAATGCGGCGGCAATGCTGT 1800
DB 1741 AATGGAGCTTGAATTCAGCAACCGCGTGGTATTCGCGCAATGCGGCGGCAATGCTGT 1800
```

Qy	1801	TCCTTTGCCAACCTTGGCGGTGATCATCTCGGTATGATCGGTTCGCTTATTTGGAAC	1860
Db	1801	TCCTTTGCCAACCTTGGCGGTGATCATCTCGGTATGATCGGTTCGCTTATTTGGAAC	1860
Qy	1861	AAATGGGACAAAGCTGCGAAAACTTGCAGCGGCACGCCCTCATCCGCGGTTCAATG	1920
Db	1861	AAATGGGACAAAGCTGCGAAAACTTGCAGCGGCACGCCCTCATCCGCGGTTCAATG	1920
Qy	1921	CGGTTATGGGTATTCGCGCGCGCGGTTCGCTTCAAAATCCGAAGCCATGCTCCACCCG	1980
Db	1921	CGGTTATGGGTATTCGCGCGCGCGGTTCGCTTCAAAATCCGAAGCCATGCTCCACCCG	1980
Qy	1981	ACCAAGTCGTCCAAATACATCTGGGACGAGCAAAATGCCAAAATCGTTACAAACCCCTGCAAA	2040
Db	1981	ACCAAGTCGTCCAAATACATCTGGGACGAGCAAAATGCCAAAATCGTTACAAACCCCTGCAAA	2040
Qy	2041	TGGCATTTGTTGGAAACACCTTTCGACGCGGAGTCAACTGCTCCATCAAGGCGCTGA	2100
Db	2041	TGGCATTTGTTGGAAACACCTTTCGACGCGGAGTCAACTGCTCCATCAAGGCGCTGA	2100
Qy	2101	CTTACCGCACAACTGCTCCGAGCATACCGCTGGGTCAACTACGTCCGAGCACGACG	2160
Db	2101	CTTACCGCACAACTGCTCCGAGCATACCGCTGGGTCAACTACGTCCGAGCACGACG	2160
Qy	2161	ACATGGCTTGGACGTTTCCGATTAAGACGCGGCATATCTGGGCATTAAGGCTACGACC	2220
Db	2161	ACATGGCTTGGACGTTTCCGATTAAGACGCGGCATATCTGGGCATTAAGGCTACGACC	2220
Qy	2221	ACGCCAAATTCCTCAACCGCTTCTCGTCAACCGTTTCGAGCGCAGCTTCGCTCGTGGCG	2280
Db	2221	ACGCCAAATTCCTCAACCGCTTCTCGTCAACCGTTTCGAGCGCAGCTTCGCTCGTGGCG	2280
Qy	2281	TACCGTTTCCAATACAAACCCAAAGCACAGGCGACTGCGCTGTACAGCGCGGCAT	2340
Db	2281	TACCGTTTCCAATACAAACCCAAAGCACAGGCGACTGCGCTGTACAGCGCGGCAT	2340
Qy	2341	TGCTCGGCTTGGCGAAGAGATCCCAACGCGTTTGACCGCATCAAACTCTTCTACAGCA	2400
Db	2341	TGCTCGGCTTGGCGAAGAGATCCCAACGCGTTTGACCGCATCAAACTCTTCTACAGCA	2400
Qy	2401	TTGCTTTAGTACCGCGGCTGCGCGTCAATTACCTTAGCGCAGAAAGTGGGTACGCTCA	2460
Db	2401	TTGCTTTAGTACCGCGGCTGCGCGTCAATTACCTTAGCGCAGAAAGTGGGTACGCTCA	2460
Qy	2461	ATGACGAGATCGGTTCGCAAGAAGCAATTAAGAGCGACAGCGTTGGGCGCACCGTCC	2520
Db	2461	ATGACGAGATCGGTTCGCAAGAAGCAATTAAGAGCGACAGCGTTGGGCGCACCGTCC	2520
Qy	2521	CGCGCTACACAGAGCCCTGTAGCGCAACGCAACGATCCGTTCGACCGCAGCGCGGCAAA	2580
Db	2521	CGCGCTACACAGAGCCCTGTAGCGCAACGCAACGATCCGTTCGACCGCAGCGCGGCAAA	2580
Qy	2581	TCTATCAGGCGCTTGGCGCATATGATTCGCGTCCGCCAAAGCAATTCGCGCTTCGACGGCG	2640
Db	2581	TCTATCAGGCGCTTGGCGCATATGATTCGCGTCCGCCAAAGCAATTCGCGCTTCGACGGCG	2640
Qy	2641	GCAGGCTGGTTACATTCACACCAACGAGCACAATCGGCTATCATCGGCAACAATG	2700
Db	2641	GCAGGCTGGTTACATTCACACCAACGAGCACAATCGGCTATCATCGGCAACAATG	2700
Qy	2701	CGCTTTTGGCATTCGGTAACTTCAGCGAATATCCGCAAAACCGTTTACCGCGCATACCGTGC	2760
Db	2701	CGCTTTTGGCATTCGGTAACTTCAGCGAATATCCGCAAAACCGTTTACCGCGCATACCGTGC	2760
Qy	2761	AAGCCATGCCCTTCAAGGCGACGACCTCATTCGTTGGGCAAAAATGTCAGGCTGATCAGG	2820
Db	2761	AAGCCATGCCCTTCAAGGCGACGACCTCATTCGTTGGGCAAAAATGTCAGGCTGATCAGG	2820
Qy	2821	ATTTTAGCGCTTCAGCCCTATCAGGTCAATGCGTTCGAAATCGCCTGACGACGCTTCCCA	2880
Db	2821	ATTTTAGCGCTTCAGCCCTATCAGGTCAATGCGTTCGAAATCGCCTGACGACGCTTCCCA	2880
Qy	2881	AATGCCCTCTGAACCGCTTTCAGACGCGCATTTGGC	2914

```

2881 AATGCGTCTGAACCGTTTCAGACGCGCATTTGG 2914

RESULT 2
US-09-815-242-6076/c
Sequence 6076, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Gene
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/491,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6076
LENGTH: 759
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(759)
US-09-815-242-6076

```

Query Match	5.8%;	Score 169;	DB 9;	Length 759;
Best Local Similarity	53.8%;	Pred. No. 1.1e-4;		
Matches 371;	Conservative 0;	Mismatches 315;	Indels 3;	Gaps 17;

QY	1	GAGTTTTCGGTTC	CGAACCGT	CATGCTT	GAGCGGAAC	ACCTGT	CCGGCAAGCGG	60
DB	695	GACGCTTTT	TACCGCACCA	AAAGTTG	CCGCGAGCCAA	TACTT	CACAGCCAGACGA	636
QY	61	CTGACCC	CCGCTTTT	TGCCCA	TCGCA	CATCGT	TAACAATCGGTTTGGTGGCAA---GCTCT	117
DB	635	GAAATTA	CGCCAGT	TTTTTGGCC	ATCGA	CATCGT	GTGATAATTGACCATCGGCATATCGCTCC	576
QY	118	TTGCGTTT	GAGCGTGG	CAGAAAGCA	AGTCAG	CACGCTT	CCGCGCTTTGGCGGCATCACC	177
DB	575	TGCATCT	CCAGGGTC	CGGCAAGCA	AGTCAG	CACATCGT	TGTACTTTTGGCGCATCAGC	516
QY	178	GCAATTTT	TGCAGATG	TCGGCGCC	GGAGTCT	CCCATCTG	TTTTCAGACGGCATACGATTCTT	237
DB	515	GCAATCTT	TAGGAATAT	CGGCGTC	GAAAGGAT	TGCATTTT	TGCGCAGACGGGCAATGATTCT	456
QY	238	TCTTGCGGCG	CGGTCGGT	TGGAAC	CTCATG	ATTG	CAGACGACGGGCGCATGCCGTTTTTTT	297
DB	455	TCGGCTTC	CGGCGGTTTT	TATGGA	GTGTCATG	TTTGGAC	TGATCATCTTTCACATCATCGCGG	396
QY	298	TEAGCAT	CGGCCAC	CGGCGCGC	CGGCGGTTT	TGCGCGGAAAAA	AGTCGATATCGGATA	357

Db 395 TGGCGTAGGCGAAGCGTTTCTTTAACTGATCATCAGCGGTAATAACTCCAGATCGATC 336
Qy 358 ATGTCGGGCGAGCGCTTTTCAATCAGCGAGTCGAGCAGTTTCAAAATAATAATATCTCCGAA 417
Db 335 ATATCAACAGGCGCGCTGTGATGCTGCACGATGAGTCAATAAAGCTCTGGTGGAA 276
Qy 418 CACGGGAAGAGCGCGCTTGGCCATGCGCTGTGAAGCTTAACAGCAGCGGCTTGTGGGC 477
Db 275 ATGCGCTGTGCGCGCTTCTTTTGGCATCTGCGAGGTAAACAGCAGCGGTTTCTGGC 216
Qy 478 AGCGGTGCGGAGCGCTTGTGCTGTGGCGCAATATCTGCGCGATGCTGCCCGCAATTC 537
Db 215 ATGCTCTCAAGAGAAATTTTCTGCCCGCATGACAGACTCCCAATTTGGAGAGGTCCGCA 156
Qy 538 AAAAATCGCGCGAATCTGACGATATCGAGGCGAGGCTTTTGAATTTGTCAGTACG 597
Db 155 TAGTGGTCAACAGCGCTTCAAGATATCAAGTCCGCTTCAAGTACGAGAGCTTCG 96
Qy 598 GCGGAAGTACGCGCGCATGCGCGCAAGAGCGGCAAGCGGCAAGTTTGGTGGCTTCGCTT 657
Db 95 GATTTCAAGCTGCGCGATATCTTTCGCGATCAGCGAGAGCATGATTTTAGGTGGCGCGTA 36
Qy 658 CGGATAAGCTTTTTCAGCTCAGGCT 686
Db 35 CCAATGACGAGATCTTTTACAGTTACGCT 7

RESULT 3

US-09-815-242-9839/c
; Sequence 9839, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9839
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(982)
US-09-815-242-9839

Query Match 5.5%; Score 161.4; DB 9; Length 882;
Best Local Similarity 53.3%; Pred. No. 3.2e-40;
Matches 364; Conservative 0; Mismatches 316; Indels 3; Gaps 1;

Qy 7 TCGCTTCCGGAACGAAAGCTGATGCTTTCAGCGGAAACCTGTCGCGCAAGCGGCTGACC 66
Db 812 TGCTTTCACGCGCCAAAGCTGCGCGGAGAGCAAAACACTTCCCTGCGCAAGCGTGAATG 753
Qy 67 GCGGCTTTTCCGCGATCGATCGTAACATCGTTTGTGGSC---AAGCTCTTTCGCT 123
Db 752 ACACCTCTTTTCCGCGATGATAGTAATACCGGAAGCTGCGGCTAATGTTCTGCTATC 693
Qy 124 TTGAGCGTGGCAGAAAGCAAGTTCAGCAGCTCTTCCGCGCTTTTTCGCGCATCAGCGCAATT 183
Db 692 TCCAGCGTGGCAGTGAAGTAACTGATTAATACATCATGCTTGTCTTTCGCGCATCAGCGCAATC 633
Qy 184 TTGAGATGTCGCGCGCGAGTCTCTCCATCTGTTTTCAGCGGATAGATTTCTTCTTTC 243
Db 632 TTGGGAATATCCGCGCGAGTCTTTCATTTTACGTAGCGCGCAAGCAATTTCTCTTCGCG 573
Qy 244 GCGCGCTGCGGTGAAGTCAATGATTCAGCAGAGGCGCGGATGCGCTTTTTCAGCA 303
Db 572 GACGCGCTGCGTGAAGTCTGATTAAGTAACTACAGCATACAGCAATAAATATGCGCATGGCA 513
Qy 304 TCCGCGCGCGCGCGCGGAGTTCGCGCGCTTTTCGCGGAAAGCTCGATATCGATAATGTCG 363
Db 512 TATCGAGCTGGCTTTTAACTCAGCATCAGCGGTAATAGCTCAAGATCGATCATATCG 453
Qy 364 GCGAGCGGCTTTCAATCAGCGAGTTCAGCAGTTCATAAATAATAATCGTCCGCAACGCGG 423
Db 452 ACCAGCGCGCTGCTGATTTGCGGCGAGTAAAGCTGAGTAACTGCTGAGTGTATTGTC 393
Qy 424 AACGAGCGCGCTTCGCGATGCGCTGCTGAAGCTTAACAGCAGCGGCTTTCGCGCGCGCG 483
Db 392 TGCTGCGCGCTTCTTTGCGCGTTCGCGGAAAGTAAACAGTAACGGAATGTGAGGATCGCA 333
Qy 484 TCGCGAGCGCTGCTGCTGTCGCGCAATCTTTCGCGGATGCTGCGCGCGCTTTCGCGGAT 543
Db 332 TCGCGGATAACAGCGCGAGCGGTAAGAAACGATTTGAGTTCGATCGATATCCATAAGTGA 273
Qy 544 TCGCGCGGAGTTCAGCATATTCAGAGGCGAGCTTTTGTATTTGCTCAAGTACGCGGAA 603
Db 272 TCCAGCGCGCTTCCAGATATCGAATGATGCTTTCGCGGAGGCGCGCTTTCGCTTTC 213
Qy 604 AGTACGCGCGCATCGCGGCGGCAAGCGGCAAGCGGATTTTGTGCGCTTCGCTTTCGATA 663
Db 212 AGCTATTGATGCTCTTCCCATCAAGCAGCATATTTTGGGCGATCCCTTCGCGAATG 153
Qy 664 ACGGTGTTTTCAGCGTACGCT 686
Db 152 ATAAGATTTTTCAGCGTACGCT 130

RESULT 4

US-09-815-242-6759/c
; Sequence 6759, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

```

RESULT 5
US-09-070-927A-275
; Sequence 275, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 275:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 6415 base pairs
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 275:
US-09-070-927A-275

Query Match 4.9%; Score 142.6; DB 10; Length 6415;
Best Local Similarity 51.5%; Pred. No. 9.9e-34;
Matches 353; Conservative 0; Mismatches 329; Indels 3; Gaps 1;

Qy 1 GAGTTTGGGTTCCGGAACCGAAGCTGATGCTTGAGCGGACACCTGTCGGCAGGCGG 60
Db 2727 GAGCGTTGCTGGGGGGAGCCAAAAGTTAGTGGCGGATCGGAACAGTGTGACCTGTTCAGCGA 2786

Qy 61 CTGACGCCGCCCTTTTTCGCCCATCGACATCGTAACTCGGTTTGGTGGCAAGCTCTTTT- 119
Db 2787 CTATCATCTCCCTAACTGTCCTCCATGACATAGTAGTAACGATGGAACGATGCAATGGG 2846

Qy 120 --CGCTTGAGCGTGGCAGAAAGAAAGTGAAGCTTTCGCGCTTTGGGGCATCAC 177
Db 2847 TACATTTTCATTTGGTTGCTGAAAGTAACGTAAGTACATCTGTGGCATCTTTGGCATCA 2906

Qy 178 GCATTTTTCAGATGTCGGCGGCCGACGTCCTCCATCTGTCTTTTCAGACGGCATACGATTTCT 237
Db 2907 GCTATTTTACAAATATCTGCTGACGCAATTTGCAATTTGGCGTAACCGCGGCAATCTCC 2966

Qy 238 TCTTGGCGCGGCTGCGGTGAACCTCATGATTGTCAGACGACGGCGGCGATGCGGTTT 297
Db 2966 TCTTGGCGCGGCTGCGGTGAACCTCATGATTGTCAGACGACGGCGGCGATGCGGTTT 297

```



```

Qy 298 TGAGCATGGCCAGCGCGCCGCGAGCGGGTTCGCCGGAAGGCTGATATCGATA 357
Db 3027 TTAGCTTCGTGAATTAACGTATCTGCCGCGAGCGGATTCGCAATTAATCAATATCCAAAT 3086
Qy 358 ATGTCGGGAGCGGCTTCAATCAGCGAGTCCGAGAGTTCMAAATAAATAATCGTCCGAA 417
Db 3087 AAATCCAGCGACCTTTTACTAATTCATGATACATGCAAAATAATTTCTCTCGAA 3146
Qy 418 CACGGGAACGAGCGGCTTCGCATCGCGTCTGAAAGTAAACAGCAGCGGCTTGTGCGGC 477
Db 3147 AAAGCCATTTCCGCGCTTCTTTTGTGTACGAAAGTTAAACATAAACGGTTTTTCCGCC 3206
Qy 478 AGCGCTGGCGAGCGTCTGGGTGCGCGCAATACTTCGCCGATGCTGCCGCGCATTC 537
Db 3207 AACGTTTCCATTAATCTGTTGGAAGATTACATACGGAATAATCAGCGACATTTTCA 3266
Qy 538 AAAAAATCGCGCGGAACTCGACGATATCGAAAGCGAGGTTTTTGTGTAAGTACG 597
Db 3267 TAATAGTCTAAAGCCATTCCACCATAATCAAAATCTAGTGTGTTGGCTCGCGCTTGTG 3326
Qy 598 GCGGAAGTAGCGCGCATCGCGGCGAGCAGCGGCGGAGGCTTGTGCGGCTT 657
Db 3327 GCCAAAATATCTCGGCTGTGGAGCAACAATGGGCAACCAATTTTGGGATTCCTTCA 3386
Qy 658 CCGATAACGGTGTGTTTGAAGGTCA 682
Db 3387 CCTATTCGCACATTTTGGACTATAA 3411

```

RESULT 6

```

US-09-815-242-7764
; Sequence 7764, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7764
; LENGTH: 3303
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(3303)
US-09-815-242-7764

```

Query Match 4.1%; Score 120.4; DB 9; Length 3303;

```

Best Local Similarity 47.9%; Pred. No. 8e-27;
Matches 463; Conservative 0; Mismatches 476; Indels 27; Gaps 3;
Qy 1307 CGCGATTTGAAGGCTTGAAGATATAAATTCCTTATTTTCHAGAGCTTGGTTGACTTA 1366
Db 111 CGCGACTTTCGCGGCTGATCGAGAGCTCGACTACATCGCGGCTTCGGCGTGAACAC 170
Qy 1367 TCTGCACCTGATGCGGCTGTTTAAATGCCCTGAAGGCAAAAGCGAGCGGCTATGCGGT 1426
Db 171 TCTCTGGCTGTGCTGCTT-----CTACCGTTCGCCACGCCGCGAGCGGTACGACAT 224
Qy 1427 CAGCAGCTACCGCGATGTCATTCGCGCACTGGGCAATAGCGGCTTTCGCGGAGTCTAT 1486
Db 225 CGCCAGTACCGTGGCTGACAGGAGCTACGCGAGCCTCGCGAGCCTCGCGCGGCTTCAT 284
Qy 1487 TCGTGGCTGTCAGAGCGCGCATTTTCGCGCTCTCGATTTTATCTTCAACACACACTTC 1546
Db 285 CGCGAGGCGCCATCGACGCGGTCTGGGGTGTATACCGAGCTGCTGATCAACCATACCTC 344
Qy 1547 CAACGAACAAGATGGCGCAACGCTGCGCCGCGCGGCGAGCCGCTTTTCGACAAATTTCTA 1606
Db 345 CGACAGCATCTGCTGTTTCATCGTCCGCGCACGCGAAGAGGATCGCGCGCGCGCA 404
Qy 1607 CTATATTTTCCCGGCGCGGATCGCGACCAATACGACCGCACCTTCGCGGAAATCTT 1666
Db 405 CTACTAGCTCTGGTGGAC---AGCGACGAGAAATACCGGGTACCGGATCATCTTCAT 461
Qy 1667 CCGGACACGACACCGCGGCGGCTTCTCGCAACTTGAAGAAGCGCTGGGTGTGAGACAC 1726
Db 462 CGACACCGAGAGTCCAACTGCGACCTTGGGACCGGTAGCCCAACAGTACTACTGGCACCG 521
Qy 1727 CTTCAATTCCTTCCATGGGACTTGATTAAGACACCGCTGGGTATTCGCGCAATGSC 1786
Db 522 CTTTATTTCCACGCGGAGCTGAACCTTCGACAAACCGCGAGGTCTCGCGGAGGTGCT 581
Qy 1787 GCGGAAATGCTTTCCTTTCGCAACTTGGGCGTTCGATCTCGCTGATGATGCGGTGTC 1846
Db 582 CGGGGTGATGCTGCTACTGCTGGACATGCGGTTCGAGGCGCTCGCGCTGAGCGGATTC 641
Qy 1847 CTTTATTTGGAACAAATGGGACAAAGCTGCGAAACCTTCGCGAGCGGCGGCGGCTCAT 1906
Db 642 CTACCTGATCGAACGCGAGCGGACCAAGCAGCAGGAGAACCTTCGCGGAGACCGGCGGCT 701
Qy 1907 SCGCGGTTCAATGCGGTTTATGCTGTTTCGCGCGCGCGCTGCTTCTTCAAAATCCGAAGC 1966
Db 702 CAAGCGCATCCGCGCGAGCTGACCGGCACTATTCGCGCGGCTGCTGCTGCGCGAGGC 761
Qy 1967 CATGCTCCACCGGACCAAGTCTGTCCTCAATACATCGG-----GAGGA 2008
Db 762 CAATCAGTGGCGGGAAGACACCGCGGCTACTTCGCGCGGAGGATGGCGGCGAGGCGGA 821
Qy 2009 CGAATGCGAAATCGGTTACAAACCGCTGCAATGGCAATTTGTTGGAACACCGCTGCGAC 2068
Db 822 CGAATGCGCAATGGCGCTTCCACTTCGCGTGGCGCGGATGATGCGGATGATGCGGATGCGCCA 881
Qy 2069 GCGCGAAGTCAACCTGCTCCATCAGGCGGTGACCTACCGGCGCAACCTGCTCCGAGCATAC 2128
Db 882 GGAGGATCGTATCCGATCACCGACATCTGCGCGAGACCGCGGACATCCCGGCGCAATTG 941
Qy 2129 CGCTGGGTCAACTAGCTGCGGAGCGACGACGATCGGCTGCGGCTGCGGCTGCGGATGAAGA 2188
Db 942 CCAATGGGCGATCTTCTGCGCAACACGACGAGCTGACCGCTGGAGATGCTACCGGACGA 1001
Qy 2189 CGCGGATATCTGGGATAGCGGCTTACGACCGCGCAATTTCTCAACCGCTTCTTCGT 2248
Db 1002 CGAGCGGACTATCTCTGGACCACTATGCGCGCGCGCGCGGCGGCTCAACCTGGG 1061
Qy 2249 CAACCG 2254
Db 1062 CATCCG 1067

```

RESULT 7

US-10-156-761-2790
 ; Sequence 2790, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 2790
 ; LENGTH: 1716
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1716)
 US-10-156-761-2790

Query Match 4.1%; Score 119; DB 14; Length 1716;
 Best Local Similarity 50.4%; Pred. No. 1.5e-26;
 Matches 352; Conservative 0; Mismatches 335; Indels 12; Gaps 2;
 QY 1304 TCCGCGGATTGAAGGCTTGAACATATAATCTTTTCAAGAGCTGGTTTGAC 1363
 DB 132 TGTGCGGACCTGAAGGCTTGAACATATAATCTTTTCAAGAGCTGGTTTGAC 191
 QY 1364 TTATCTGCACCTGATCGCGCTGTTAAATGCTTGAAGCAAAAGCGAGCGGCTATGC 1423
 DB 192 CTGCTGTGCTGCGCGCTGTTTCAAGTCCCTC-----CTGCGGAGCGGCTTACGA 245
 QY 1424 GGTACAGCTACCGGATGTCATCGGCACTGGGCAATAGGCGACTTGGCGAAGT 1483
 DB 246 CGTCTCGGACTACCGCGGCTGTCGCGGAGTTGCGTGACCTGGCGGACTTGTGCGAGTT 305
 QY 1484 CATCTGCTGCGCTGACGAGCGGCTTTCGCGGCTGCTGCAATTTTATCTTCAACACAC 1543
 DB 306 CGTGGAGCGGCGGCGGCGGCGGCTGCGGCTGCTGCAATTTTATCTTCAACACAC 365
 QY 1544 CTCCAGCAGACAGAAATGGGCGCAACGCTGCGCGCGCGGCGGCGGCTTTCGCAATTT 1603
 DB 366 CAGCGAGCTGACCGCGGCTTCCAGGAGTCCCGAGGAGTCCCGAGGAGTCCCGAGGAGT 425
 QY 1604 CTACTATTTTCCCGAGCGCGGATGCGCGGAGTCCCGAGGAGTCCCGAGGAGTCCCGAGGAGT 1663
 DB 426 CTACTATTTTCCCGAGCGCGGATGCGCGGAGTCCCGAGGAGTCCCGAGGAGTCCCGAGGAGT 481
 QY 1664 CTTCGCGGAGCGGCGGCTTTCGCGGAGTCCCGAGGAGTCCCGAGGAGTCCCGAGGAGT 1723
 DB 482 --TCGACAGCGGCGGCTTTCGCGGAGTCCCGAGGAGTCCCGAGGAGTCCCGAGGAGT 539
 QY 1724 GACCTTCAATTTCTTCAATGGGAGTGAATTAAGCAACCGCGGCTATTTCCGCGCAAT 1783
 DB 540 CCGCTTCTTCTCCAGCGGCGGCTTCAATTAAGCAACCGCGGCTATTTCCGCGCAAT 599
 QY 1784 GCGGCGGAGGAGTGTCTTCTTCAATTTGGGCGGTTGACATCTGCGTATGATGCGGT 1843
 DB 600 CGTCTCCGCGGCTTCTTCTTCAATTTGGGCGGTTGACATCTGCGTATGATGCGGT 659
 QY 1844 TGCCTTTATTTGGGAGCAAAATGGGAGCAAGTGGGAGCACTGGGAGCACTGGGAGCACTGGGAGCACT 1903
 DB 660 GCGGTACCTGTACAGCAGGAGGAGCACTGGGAGCACTGGGAGCACTGGGAGCACTGGGAGCACT 719

QY 1904 CATCGCGGCTTCAATGGCTTATGCGGCTTATGCGGCGCGCGGCTTCTTCAAAATCGA 1963
 DB 720 CCTGAAGCGGCTGCGAAGGAGATCGACAGCACTATTCGAGACAGCTGCTGCTCGCGGA 779
 QY 1964 AGCCATGCTCCACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2002
 DB 780 GGCGAACCCAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 818
 RESULT 8
 US-10-156-761-1
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(9025608)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Query Match 4.1%; Score 119; DB 14; Length 9025608;
 Best Local Similarity 50.4%; Pred. No. 2.1e-24;
 Matches 352; Conservative 0; Mismatches 335; Indels 12; Gaps 2;
 QY 1304 TCCGCGGATTGAAGGCTTGAAGATATAATCTTTTCAAGAGCTGGTTTGAC 1363
 DB 3441669 TGTGCGGACCTGAAGGCTTGAAGATATAATCTTTTCAAGAGCTGGTTTGAC 3441728
 QY 1364 TTATCTGCACCTGATGCGCTGTTAAATGCTTGAAGCAAAAGCGAGCGGCTATGC 1423
 DB 3441729 CTGCTGTGCTGCGCGCTGTTTCAAGTCCCTC-----CTGCGGAGCGGCTTACGA 3441782
 QY 1424 GGTACAGCTACCGGATGTCATTCGCGCACTGGGCAATAGGCGACTTGGCGAAGT 1483
 DB 3441783 CGTCTCGGACTACACCGGCTGCTGCGGAGTTCGCTGACCTGGCGGAGTTCGCTGAGTT 3441842
 QY 1484 CATCTGCTGCTGACGAGCGGCTTTCGCGGCTGCTGCAATTTTATCTTCAACACACAC 1543
 DB 3441843 CGTGGAGCGGCGGCGGCTTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGT 3441902
 QY 1544 CTTCAGCAACAGAAATGGGCGCAACGCTGCGCGCGCGGCGGCGGCTTTCGCAATTT 1603
 DB 3441903 CAGCGAGCTGACCGCGGCTTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGT 3441962
 QY 1604 CTACTATTTTCCCGAGCGGCGGATGCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGT 1663
 DB 3441963 CTACTATTTTCCCGAGCGGCGGATGCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGT 3442018
 QY 1664 CTTCGCGGAGCGGCGGCTTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGT 3442076
 DB 3442019 --TCGACAGCGGCGGCTTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGTTCGCGGAGT 3442076
 QY 1724 GACCTTCAATTTCTTCAATGGGAGTGAATTAAGCAACCGCGGCTATTTCCGCGCAAT 1783

Db 3442077 CGCTTCTTCTCCACAGCCGACCTCAACTACGAGAACCGCGGTGCAGGAGAGAT 3442136
QY 1784 GCGGCGGAATGCTGTTCTTCCCACTTGGCGGTGACATCTCGGTATGATCGGT 1843
Db 3442137 CGTCTCGGCTTGGGTCTTGGCTGACCTCGCATCGAGGTTCCGCTCGACCGGT 3442196
QY 1844 TGCTTTATTTGAAACAAATGGGCAAGCTGCGAAACCTGCGGACGCGCCCT 1903
Db 3442197 GCGTACCTGTACAGCAGGAGGACCAACTGCGAGAACCTGCGGCGACGCGGTT 3442256
QY 1904 CATCCGCGGTCAATGCGGTATGCGGTATGCGGCGCGCGGTCTTCTCAATCCGA 1963
Db 3442257 CTTGAAGCGGTGCGGCAAGAGATGACACGCACTATCCGACACGCTGCTGCGGA 3442316
QY 1964 AGCATCTGTCACCCCGACCAAGCTGTCCTCAATACATCGG 2002
Db 3442317 GCGGACCACTGCGCGGAGGAGCTGCTGACTACTTCGG 3442355

RESULT 9

US-09-974-300-399/c

; Sequence 399, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.

; APPLICANT: Clausen, Ib Groth

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; TITLE OF INVENTION: Expression

; FILE REFERENCE: 10085.500-US

; CURRENT APPLICATION NUMBER: US/09/974,300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279,526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 399

; TYPE: DNA

; LENGTH: 483

; ORGANISM: Bacillus licheniformis

US-09-974-300-399

Query Match 4.0%; Score 118; DB 10; Length 483;

Best Local Similarity 57.0%; Pred. No. 1,6e-26;

Matches 236; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY 1 GAGTTTTCGTTCCGAAACCGAAGCTGATGCTGAGCGCGAAGACCTGTCGCGCAAGCGG 60
Db 424 GAAGCTTCTTTTGGCGCACCGAAGCTGATGCTGTAACCGAATACCTGCGCGCGCGG 365
QY 61 CTGACCGCCCTTTTGGCCCATCGACATGTAACATCGTTTGGTGGCA---AGCTCT 117
Db 364 CTGATGATACCTTTGCGCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 305
QY 118 TTGCTTTGAGCGTGGCAGAAAGAAAGTACGACGCTTTCGCGCTTTTCGCGCATCAC 177
Db 304 TTGCTGATGAGTGGCTTCAAGCAGACGCAAGACATCAGCGGTGTTTCGCGCATCAC 245
QY 178 GCAATTTTGCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAAGCGCATGATTTCT 237
Db 244 GCGATTTTGCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAAGCGCATGATTTCT 185
QY 238 TCTTGGCGCGCGTGGCTGAAATCTCATGATGTCAGAGCAGGCGCGGATGCGGTTTTT 297
Db 184 TCTGACGCTGGTCTTTTCAAAATCATGATGATGATGATGATGATGATGATGATGATGAT 125
QY 298 TGACGATGCGCACGCGCGCGCGAGCGGCTTTCGCGGAAAGAAAGCTGATGATGATA 357
Db 124 TGTGCTGTTTCCACGACGCGCTTTCACCTTGTCTTTCGCTGATGATGATGATGATGAT 65
QY 358 ATGTGCGGCGCGGCTTTCAATCAGCGAGTGCAGAGTTCAAAATAATAATG 411

Db 64 ACATGCGGACCGCTGTTTAAACACCGCTTCATTTAAGAAATAATATCGG 11
RESULT 10
US-09-738-626-2527
; Sequence 2527, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 2527
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2527

Query Match 3.8%; Score 111.8; DB 10; Length 1794;

Best Local Similarity 48.3%; Pred. No. 3.1e-24;

Matches 422; Conservative 0; Mismatches 427; Indels 24; Gaps 3;

QY 1305 GCGGCGATTTCGAAGGCTTGAAGATAAAATTCCTATTTCAGAGACTTGGTTGACT 1364
Db 220 GTGCGATCGTTGAAGGCTTGACCGAAATATGATATCATCCAGTGGCTGGGCTGGAT 279
QY 1365 TATCTGCACCTGATGCGCTGTTTAAATGCTTGAAGGCAAAAGCGAGCGGCTATGCG 1424
Db 280 TGCATTGGATCCACCGCTTTTATGATTCCCC-----ACTGCGGACGCGGTTACGAT 333
QY 1425 GTCAGCAGCTACGCGATGTCATCGGCACCTAGGCGACTTGGCGGAAGTC 1484
Db 334 ATCCGCACTTCGCGAATCTGCGGAAATTCGCGACCGTGCATGATGACTTCGTGGAATC 393
QY 1485 ATTGCTGCGCTGCAGAAACCGGCAATTCGCGCGTGTGATTTATCTTCAACCAACC 1544
Db 394 GTTGACCAACGCGCACCGCGCTGCTGCTTATATCCGACTTGGTCAATGAATCACACC 453
QY 1545 TCCAGCAACAGATGCGCGCAACGCTGCGCGCGCGGACCGCTTTTCGACAAATTC 1604
Db 454 TCGACCAACAGATGCTTCCAAAGATTCGCGCGCAACCCACCGCCCTTACGAGAT 513
QY 1605 TACTATATTTTCCCGACCGCGGATGCGGCAATACAGCAACCGCTGCGCGAAATC 1664
Db 514 TTTATGTTGGAGGATGATCCACCTGTATACAGGAGCGGATCAT-----CTTT 567
QY 1665 TTCCCGCAACAGACCGCGGCTTTCGAACTGGAAGAGAGCGCTGGGTGTGACG 1724
Db 568 GTAGATACAGAGAAATCCAACTGGACCTATGATCGGTGCGGCGAGTCTTCTGGCAC 627
QY 1725 ACCTTCAATTCCTTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCGCGCAATG 1784
Db 628 CGTTTCTTCCACCAACCAAGCTCACTAGCAACCCCGCACTTCCAAGAGGCGCATG 687

Qy	1785	CGCGCGGAATGCTGGTTCTCTTGCACAACTTGGCGGCTTGACATCTCTGCGTATGGAATGCGGTT	1844
Db	688	CTAGATGCTTTCGGCTTTCTGCTGGAACTTGGGACTTGTATGGTTTCCGACTAGATGCGGTT	747
Qy	1845	GCCTTTATTTGGAAACAAATGGGGACAAGCTTGGGAAACCTTGGCGAGGCGCACGCCCTC	1904
Db	748	CTTTATCTTTTGHACCGGAAGGACCAACGGCGGAAACCTTCAAGAAACCCAGATTTC	807
Qy	1905	ATCGCGCGTTCAATGCGGTTATGCGCGCGCCGCGTGTTCCTTCAAATCCGAA	1964
Db	808	CTCAAACTGTGTGCTCTGTTCATTGAGAGGAATACCCGCGCGAATCTCTGCTCGCAGAA	867
Qy	1965	GCATCTGTCCACCCCGACCAAGTGTGCTCAATACATCGGGCA-----GGACGAA	2012
Db	868	GCCAAACAAATGGCCCCCAAGATGTGTGCAATCTCTGCGTGGAAAGAACAAAGCGCATGAA	927
Qy	2013	TGCCAAATCGGTTACAAACCCCTGCAAAATGGCATTTGTGTGGAAACACCTTGGCACGCGC	2072
Db	928	TGCCACATGCGCTTCGATCTCCCTTTGTATGCGCGCGATCTTCTATGGGAGTTCCGCAAGGT	987
Qy	2073	GAAGTCAACCTGTCTCATAGGCGCTGACCTACCGCCACAACCTTGCCTCGGACATACCGCC	2132
Db	988	TACGCAACCCCGATCAGTGAATCTCTGGCAACACCCCGGAGATTCCTCCAGACTGCCCAA	1047
Qy	2133	TGGGTCAACTACGTCCGACGCGACGACATC	2165
Db	1048	TGGGGTATTTTCTGGGTAAATCATGATGAGCTC	1080

RESULT 11

US-09-738-626-1

: Sequence 1, Application US/09738626

Publication No. US20020197605A1

: GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

: APPLICANT: ANDO. SEIKO

APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MITSUO

APPLICANT: HAYASHI, MITSUO
APPLICANT: OCHIAI, KETKAI

APPLICANT: UCHIAL, KEITH
APPLICANT: YOKOT, HARITH

; APPLICANT: YUKOI, HARUHI
; APPLICANT: TATEISHI, NAOMI
; APPLICANT: TATEISHI, NAOMI

APPLICANT: TATEISHI, NA
APPLICANT: SENO, AKIHIRO

); APPLICANT: SENOH, AKIHI
: APPLICANT: Takeda MASAT

APPLICANT: IZDA, MASAT
APPLICANT: OZAKT AKTO

APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYACETATES

; TITLE OF INVENTION: NOVEL
 ; PRI REFERENCE: 349-12E

PILE REFERENCE: 249-125

;	CURRENT APPLICATION NUMBER:	US/0
:	CURRENT FILING DATE:	2000-12-18

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 1988-100000

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: 3

; PRIOR FILING DATE: 2000-08-

; NUMBER OF SEQ ID NOS: 7059

```
; SOFTWARE: PatentIn ver. 3.0
```

; SEQ ID NO 1

; LENGTH: 3309400

TYPE: DNA

; ORGANISM: Co

Query Match 3.8%: Score 111.8: DB 10: Length 3309400:

Query Match	Score	DB ID	Length
Best Local Similarity	48.3%		
Pred. No. 2.3e-22:			

```

Best local similarity 48.3%, Freq: NO: 2:38-22,
Matches 422: Conservative 0: Mismatches 427: Indels 24: Gaps 3:

```

1305 GCGGGCGATTGTGAAGGGCTTGAAGATATAATTCCTTAGTTTCAAGAGCGTTGGTTGACT 1364

[illegible]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

[illegible]

```

: PRIOR FILING DATE: 2001-08-02
:
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 7374
:
: LENGTH: 1749
:
: TYPE: DNA
:
: ORGANISM: Streptomyces averm
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1749)
: US-10-156-761-7374

```

Query Match	3.8%	Score 111:	DB 14:	Length 1749:
Best Local Similarity	49.6%	Pred. No. 5.5e-24:		
Matches 347:	Conservative	0:	Mismatches 340:	Indels 12: Gaps 2:
QY	1304	TGCGCGCATTTGGAAGGGCTTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTTGAC	1363	
DB	132	TGTCGGGACCTTGAAGGGCTTGACCGCCAAAGCTGACTATCTGCAGTGGCTGGCGTGGGA	191	
QY	1364	TTATCTGCACCTGATGCGCGCTGTTTAAATGCGCTTGAAAGGCAAAAGCGACGGCGCTATGC	1423	
DB	192	CTGCGTGTGCTGCGCGCGTTCCTTCAGTCCCC-----CCTGGCGACCGCGCGCTACGA	245	
QY	1424	GCTCAGCAGCTTACCGCGATGTCATTCGCGCACTGGGCACAATAGGCGACTTGC CGCGAAGT	1483	
DB	246	CGTCTCGGACTACACCGCGGTGTCGCCGAGTTCGGTGACCTGCGCGACTTCGTCGAGTT	305	
QY	1484	CATTGTGCGCTGGAAGGAGCGGCGATTTGCGCGTGTGTCGATTTTATCTTCAACACAC	1543	
DB	306	CGTGGACGCGCGGCCACCAGCGCGGCATTCGCGTGATCATCGACTTCGTGATGAACCCAC	365	
QY	1544	CTCCACGGAACACGAATGGCGGCAACGCTGCGCGCGCGGACCGCGTTTTCGACAAATTT	1603	
DB	366	CAGTGACCTGCATCCGTGGTTCAGGAGTCCCGCAGCAACCCGACGCGCCCTACGCGGA	425	
QY	1604	CTACTATATTTTCCCGACCGCGGATGCGCGACCAATAGACCGCGACCCCTGCGCGAAAT	1663	
DB	426	CTACTAGCTGTGGCGCCAGCAGCAAGCAATACAGAGCCCGCGGATCATCTTCG----	481	
QY	1664	CTTCCCGACACGACCAACCGCGCGGCTTCTTCGCAACTGGAAGACGGAAGCTGGGTGTGGAC	1723	
DB	482	--TGACACGAGGCTTCCAACTGGACGCTACGACCGGTCGCGAAGCAGTACTACTGGCA	539	
QY	1724	GACCTTCAAATTCCTTCAAATGGGACTTGAATATGAGCAACCCGTGGGATTCCTCGCGCAAT	1783	
DB	540	CGGCTTCTTCTCCACACGCGCGACCTCAACTACGAGAGTCCGCGCTCCAGGAGGAGAT	599	
QY	1784	GGCGGGCGAATGTGTGTCTTTCGCACTTCGGCGTTTGACATCTCGGTATGGATGCGGT	1843	
DB	600	CCTGGCGGCGCTGGGTTCTGGCTCGATCTGGGATATCGACGGCTTCAGGCTGGAGCGCGT	659	
QY	1844	TGCCTTTATTTGGAAACAAATGGGGACAAGCTGCGAAACCTGCGCAGGCGGCGACGCCCT	1903	
DB	660	CCCCTACTGTATCAACAAGAGGGGACGAACTGGCGAGAACCTGCGCGCGACGCGACGAGTT	719	
QY	1904	CATCCGCGCGTTCAATGCGGTTATGCGGTATTGCGCGCGCGCGGTGTTCTTCAAAATCGA	1963	
DB	720	CCTGAAGGGGTGGCAAGGAGATGCAACGCACTATCCGACACAGCTGCTGCTCGCGGA	779	
QY	1964	AGCCATGCTCCACCCCGACCAAGTGTCCAAATACATCGG	2002	
DB	780	GGCGAACCAAGTGGCGGAGGACGCTGGTCTGCACTACTTCGG	818	

RESULT 13

RESULT 13	
US-10-156-761-1713	
; Sequence 1713, Application US/10156761	
; Publication No. US20030119018A1	
; GENERAL INFORMATION:	
; APPLICANT: OMURA, SATOSHI	
; APPLICANT: IKEDA, HARUO	
; APPLICANT: ISHIKAWA, JUN	
; APPLICANT: HORIKAWA, HIROSHI	
1992 CAATACATCGGCAGCAGCAATGCCAAATCGGTTTACAAACCCCTGCAAAATGCAATGTTG 2051	Qy
799 CCGACACCGAAGCTTCGCGGCTATCTGCGCCGACGAACTGCACACGCTTCAAC 858	Db
2052 TGGACACCCCTGCCACGCGGAGTCAACTCTCCATCAGGCGCTGACCTAC 2105	Qy
859 TTCTCTTCTCTCTGTCCTCGGACGCGCGGGGCGCTGCGCAGGCCATCGAC 912	Db

Search completed: November 8, 2003, 20:14:57
Job time : 917 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 12:46:16 ; Search time 5415 Seconds

(without alignments)
13079.085 Million cell updates/sec

Title: US-09-843-007A-1

Perfect score: 2914

Sequence: 1 GAGTTTCGGTCCCGAAC.....CGTTTCAGCGCATTTGCG 2914

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pin.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rtd.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	104	3.6	1055	28	AF029476	AF029476
2	92.6	3.2	518	9	AI964123	AI964123
3	90.4	3.1	785	29	B2578463	B2578463
C 4	89.4	3.1	968	29	B2554240	B2554240

C 5	84.8	2.9	527	9	AI964122	AI964122
C 6	73.2	2.5	606	10	BE249225	BE249225
C 7	57	2.0	367	13	BU003396	BU003396
C 8	54.6	1.9	763	14	CB997861	CB997861
C 9	54.6	1.9	845	14	CB996585	CB996585
C 10	53	1.8	912	11	CNS08POT	CNS08POT
C 11	50.4	1.7	1201	13	EX381961	EX381961
C 12	49.8	1.7	659	9	AI107032	AI107032
C 13	49.8	1.7	703	12	BMS87500	BMS87500
C 14	49.8	1.7	1981	28	BH770895	BH770895
C 15	49.2	1.7	551	9	AI386572	AI386572
C 16	49.2	1.7	589	9	AI517021	AI517021
C 17	49.2	1.7	589	12	BI366411	BI366411
C 18	49.2	1.7	603	12	BI372255	BI372255
C 19	49.2	1.7	604	12	BI228916	BI228916
C 20	49.2	1.7	607	12	BI579016	BI579016
C 21	49.2	1.7	654	12	BI173333	BI173333
C 22	49	1.7	884	28	BH149112	BH149112
C 23	48.4	1.7	925	28	BH132691	BH132691
C 24	47.8	1.6	638	9	AI402332	AI402332
C 25	47.8	1.6	641	9	AI292849	AI292849
C 26	47.4	1.6	857	28	AZ530946	AZ530946
C 27	46.2	1.6	644	14	CA631367	CA631367
C 28	46	1.6	527	12	BI628794	BI628794
C 29	46	1.6	531	12	BI566079	BI566079
C 30	46	1.6	534	9	AA803956	AA803956
C 31	46	1.6	568	9	AI107511	AI107511
C 32	46	1.6	570	12	BI566486	BI566486
C 33	46	1.6	590	12	BI586057	BI586057
C 34	46	1.6	604	9	AI238307	AI238307
C 35	46	1.6	615	12	BI587736	BI587736
C 36	46	1.6	616	12	BI564072	BI564072
C 37	46	1.6	621	12	BI634521	BI634521
C 38	46	1.6	623	9	AI389773	AI389773
C 39	46	1.6	626	9	AI297229	AI297229
C 40	46	1.6	637	12	BI241727	BI241727
C 41	46	1.6	643	9	AI063787	AI063787
C 42	46	1.6	648	12	BI583466	BI583466
C 43	45.6	1.6	645	14	CA559056	CA559056
C 44	45.2	1.6	497	29	CC169594	CC169594
C 45						

ALIGNMENTS

RESULT 1
AF029476/c
LOCUS AF029476 1055 bp DNA linear GSS 29-AUG-2000
DEFINITION typhimurium genomic clone 472-13, genomic survey sequence.
ACCESSION AF029476
VERSION AF029476.1 GI:2571006
KEYWORDS GSS.
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
REFERENCE 1 (bases 1 to 1055)
AUTHORS Wong,R.M.-Y., Wong,K.K., Senson,N.R. and McClelland,M.
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library: comparison to the Escherichia coli K12 genome
JOURNAL FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
MEDLINE 99243757
PUBMED 10227170
COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3059 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifsci.sdu.edu
Class: Shotgun.
Location/Qualifiers
source
1. .1055

AF029476 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 472-13, genomic survey sequence.

```
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="LT2"
/db_xref="taxon:602"
/clone="472-R3"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 277 a 253 c 250 g 258 t 17 others
ORIGIN
Query Match 3.8%; Score 104; DB 28; Length 1055;
Best Local Similarity 58.3%; Pred. No. 1.5e-17;
Matches 201; Conservative 0; Mismatches 141; Indels 3; Gaps 1;
QY 1 GAGTTTGGCTTCCCGAAGCGAAGCGTATGCTTGAGCGGCAACACCTGTTCGGCAAGCGG 60
DB 379 GATGCTTTTTCACCGCGCAACAGTTCGGCAGAACGACACTTCGCGCGCAGACGA 320
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGAATCAATCGTTTGTGGCAA---GCTCT 117
DB 319 GATATTACCCCGTTTTCGACATCGACATGTAATATCGGACATNCGCATAGCGCTCC 260
QY 118 TTGCGTTTCAGCGTGCAGAAACAAAGTCAGACGCTTCGCGCTTTCGGCATCAC 177
DB 259 TGCATTTTCAGTGGCGGTAAATAGGTTCAGACATCGGCTTTAGTCTGTGGCATGCG 200
QY 178 GCAATTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGGCATACGATTTCT 237
DB 199 GGCATCTTCGGAATATCAGCGCCAGTCTCTCATTTTCAGCAGACGCTGACATCTCT 140
QY 238 TCTTGGCGGGCGGCGGTGAATCATGATTCGACGACGCGCGCGATCGCGTTT 297
DB 139 TCCGCTGCGGGCGTTTATGAAATCATGTTAGACATGATCATCCGCAATGTGTGA 80
QY 298 TGAGCATGCGCCACGCGCGCGCGAGCGGCTTTCGCGGAAAAA 342
DB 79 TGAGCATAGCCGAGCGTGGCTTTCACCTCATCGTCGCGGTAAAA 35
RESULT 2
LOCUS AI964123 518 bp mRNA linear EST 20-AUG-1999
DEFINITION BST269237 zebrafish, Steve Ekker Danio rerio cDNA clone RZBA17,
mRNA sequence.
ACCESSION AI964123 GI:5756836
VERSION AI964123.1
KEYWORDS Danio rerio (zebrafish)
SOURCE Danio rerio
ORGANISM Danio rerio
REFERENCE 1 (bases 1 to 518)
AUTHORS Quackenbush, J., Lee, N., Tomb, J.-P. and Ekker, S.
TITLE Generation of ESTs from zebrafish
JOURNAL Unpublished
COMMENT Other ESTs: EST269236
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnc@tigr.org.
FEATURES
source
1..518
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="RZBA17"
/tissue_type="whole body"
/dev_stage="maternal"
/lab_host="XLI-Blue MRF"
```

```
/clone_lib="zebrafish, Steve Ekker"
/note="Vector: pAD-GAL4; Site 1: EcoRI; Site 2: XhoI; This
library was constructed by directed cloning into the
EcoRI-XhoI sites of pAD-GAL4. Estimated Average insert size
approximately 1.2kb based on 15 clones, estimated titer of
excised phagemid library = 3.78 x 10^6 cfu/ml"
BASE COUNT 127 a 121 c 123 g 141 t
ORIGIN
Query Match 3.2%; Score 92.6; DB 9; Length 518;
Best Local Similarity 56.4%; Pred. No. 1.6e-14;
Matches 189; Conservative 0; Mismatches 143; Indels 3; Gaps 1;
QY 1 GAGTTTTCGTTCCCGAAGCGAAGTGTGATGCTTGAGCGGCAACACCTGTTCGGCAAGCGG 60
DB 133 GACGCTTTTTCACCGCACCAAAAGTTCCCGCGGAGCAATATCTTCACCGCCAGACGA 192
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAAATCGGTTTGGTGGCAA---GCTCT 117
DB 193 GAAATAGCGCAGTTTTTGGCCATCGACATCGTATATGACGATCGGCATCTGCTCC 252
QY 118 TTGCGTTTTCAGCGTGGCAGAAAGCAAGTCAGACGTTTCGCGCTTTCGGCGCATACC 177
DB 253 TGCATCTCCAGGTTCGCGGCAAGCAACGTCAGACATCGCTGTGACTTTTCGCGCATCAGC 312
QY 178 GCAATTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGCGCATACGATTTCT 237
DB 313 GCAATCTTAGGAATATCGCGTTCGAGGATTCGATNTTCGCGCAAGCAATGATTTCT 372
QY 238 TCTTTCGCGCGCGTTCGCGTGAACCTCATGATTCAGACGCGCGCGCATGCGTTT 297
DB 373 TCGGCTTCGCGGTTTATGGAAGTCATGTTGACATGACTTTCATCATCATCGCGG 432
QY 298 TGAGCATGCGCACGCGCGCGCGCGCGGCTTC 332
DB 433 TGGCGGTAGGCAACGCGGCTCANNCCGCTGCGACTC 467
RESULT 3
LOCUS BZ578463 785 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_5885.x1 msh Pseudomonas aeruginosa genomic clone msh2_5885,
genomic survey sequence.
ACCESSION BZ578463
VERSION BZ578463.1
KEYWORDS GI:27213524
SOURCE GSS
ORGANISM Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 785)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) in press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source
1..785
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_5885"
/dev_stage="maternal"
/lab_host="XLI-Blue MRF"
/note="Environmental isolate. Whole genomic shotgun"
```

```

BASE COUNT      154 a      274 c      224 g      131 t      2 others
ORIGIN

Query Match      3.1%; Score 90.4; DB 29; Length 785;
Best Local Similarity 49.3%; Pred. No. 8.1e-14;
Matches 266; Conservative 0; Mismatches 271; Indels 3; Gaps 1;

QY 1476 CGGAACTCAATGTGGTGGTGCACGAAGCCGGCATTTCCCGCGTCGTCGATTTTATCTTC 1535
DB 90 CGCGGTTTCATCGCGGAGGCCATCGACGGGTCTGCGGGGATTTACCGAGCTGGTGATC 149

QY 1536 AACGACACTCCACGACGACGAATGGCGCAAGCTGGCGCGCGCGCGCGCGCGCTTTTC 1595
DB 150 AACGACACTCCGACGACGATCCCTGGTTTCATCCGCGCGCGCGCGCGCGCGGATCG 209

QY 1596 GACAAATTTCTACTATATTTTCCCGACCGCGCGATGCCCAATACGACCGCACCTG 1655
DB 210 CGCGCCCGGACTA--CTAGCTTGGTGGACAGCGAGGAAATACGAGGTACCGG 266

QY 1656 CGGAAATCTTCCCGACGACGACCGCGCGCTTCTCGCAACTGGAAGACGACGCTGG 1715
DB 267 ATCATCTTTCATCGACACCGGACGATCCAACTGGACCTGGGACCGGTCGCCCAACAGTAC 326

QY 1716 GTGTGGAGGACCTTCAATCTCTTCCATGGGACTTGAATTACGACACCGCTGGGTATTC 1775
DB 327 TACTGGACCGCTTCTATTTCCACCGCGGACCTGAACTTTCGACACCGCGAGGTCCG 386

QY 1776 CGGCAATGGCGGCGGAAATGCTGTTCTTCCCAACTTGGCGGTGACATCTCGCGTATG 1835
DB 387 CGGAGGTGTCGGGGTGATGGCTACTGCTGGTGGACATGGGCTCGACGCGCTGGCGCTG 446

QY 1836 GATGGGTTGGCTTTATTTGGAAACAAATGGGACGACCTCGGAAACCTGGCGAGG 1895
DB 447 GACGCGATTCCCTACCTGATCGAACCGGACGCGACCGGACGAGAGAACTTCCGCGAGAC 506

QY 1896 CACGCGCTCATCGCGCGGTTCATCGCGTTATGGTATTTGCGCGCGCGCGCTTCTTC 1955
DB 507 CACGAGTGTCTCAGCGCATCGCGCGCGAGCTGGAAGCGGCACTATCCGACCGCATCTG 566

QY 1956 AATTCGAGGACATGCTTCACCGCGACCAAGTGTCCAAATACATTCGCGGAGGACGAATGC 2015
DB 567 CTGGCGGAGGACCAACAGTGGCGCGGAGACACCGCGCGCTACTTTCGCGTGGAGGATGC 626

RESULT 4
BZ554240/c
LOCUS      BZ554240      968 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION pacs1-60_4587.y2 pacs1-60 Pseudomonas aeruginosa genomic clone
ACCESSION BZ554240
VERSION   BZ554240.1 GI:27161328
KEYWORDS  GSS.
SOURCE    Pseudomonas aeruginosa
ORGANISM  Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 968)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.B., Hastings,M.,
          Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol., (2002) In press
COMMENT   Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.
          Location/Qualifiers
            source
              1. .968

```

```

/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone_pacsi="pacs1-60_4587"
/clone_lib="pacs1-60"
/notes="clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT      183 a      285 c      272 g      228 t
ORIGIN

Query Match      3.1%; Score 89.4; DB 29; Length 968;
Best Local Similarity 51.4%; Pred. No. 1.7e-13;
Matches 207; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY 1634 CGACCAATACGACCGCACCGCTGGCGAAATCTTCCCGACCGACCGCGCGCTTCTC 1693
DB 520 CGAAGATACCGGGTACCGGATCATCTTCATCGACCGGACGCTCCAACTGGACCTG 461

QY 1694 GCACTGGAAGACGAGCGCTGGGTGGACGACCTTCAATCTTCTTCAATGGGACTTGA 1753
DB 460 GGACCGGTGGCGCCCATCACTACTTGGCACCGCTTCTATTCACGACCGGACCTGAA 401

QY 1754 TTACACCAACCGGTGGGTATTCGCGCAATGGCGGGAATGCTGTTCTTCCCACTT 1813
DB 400 CTTGCTCAACCGCGAGTCTCGCGGAGGTCTCGGGGTGATCGCTACTGGCTGGACAT 341

QY 1814 GGGCGTTGACATCTCGTATGATCGGTTGCGCTTTATTTGGAAACAAATGGGACAAG 1873
DB 340 GGGCGTGGACGGCTCGCGCTGTACGCGATTCCTACTGATCGAACGGGTGCGCACCG 281

QY 1874 CTGGGAACCTGCGCAGCGCGCGCTCATCCGCGGTCAATGCGGTATGCGTAT 1933
DB 280 CAGCGAAGACCTGCGCGGAGACCCACGAGTGTCTCAAGCGCATCCGCGCGAGCTGGA 221

QY 1934 TGCGCGCGCGCGCTGTTCTTCAATCCGAAGCATCGTCCACCGCGACCAAGTCTGCA 1993
DB 220 GCACATATCGACCGCATGCTGCTGTCGCGACCGCAACGAGTGGCGGAGACACCGCT 161

QY 1994 ATACATCGGGGAGGACGAATGCAAACTGTTTACACACCGCTG 2036
DB 160 GTACTTCTGCGGAGCATGACGACGAGCGGAGCAATCCCTG 118

RESULT 5
AI964122/c
LOCUS      AI964122      527 bp      mRNA      linear      EST 20-AUG-1999
DEFINITION EST269236 zebrafish, Steve Ekker Danio rerio cDNA clone RZBAAL7,
          mRNA sequence.
ACCESSION AI964122
VERSION   AI964122.1 GI:5756835
KEYWORDS  EST.
SOURCE    Danio rerio (zebrafish)
ORGANISM  Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
          ; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 527)
AUTHORS   Quackenbush,J., Lee,N., Tomb,J.-F. and Ekker,S.
TITLE     Generation of ESTs from zebrafish
JOURNAL   Unpublished
COMMENT   Other_ESTs: EST269237
          Contact: John Quackenbush
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 3528
          Fax: 301 838 0208
          Email: johnq@igr.org
          3 prime sequence.
          Location/Qualifiers
            source
              1. .527
              /organism="Danio rerio"
              /mol_type="mRNA"

```


http://compugenomics.ucdavis.edu/
 Unpublished
 Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozikeatgc.org [michelmore@vegmil.ucdavis.edu]
 singleton see http://cspdb.ucdavis.edu/ for details.
 Plate: QG35 row: B column: 11.

FEATURES source

1. 367
 /location/Qualifiers
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cultivar="L. serriola"
 /db_xref="taxon:4236"
 /clone="QG35B11"
 /lab_host="E.coli"
 /note="Vector: pGEGHU lettuce serriola"
 /note="Vector: pBRCDNA5f1AB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cspdb.ucdavis.edu/TAG_SEQ=Not found"
 TAG_SEQ=Not found"
 75 a 115 c 126 g 51 t

BASE COUNT ORIGIN

Query Match 2.0%; Score 57; DB 13; Length 367;
 Best Local Similarity 56.8%; Pred. No. 0.00013;
 Matches 105; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1414 GGGCTATCGGTACACGACCTACCGGATGTCATCGGCACTGGCCAAATAGGGACT 1473
 DB 1 GGGCTGGGAGGTGATCGAGTAGTACGCAACGGGACCCACGTCTGGGCACGCTGAGGAT 60

QY 1474 TGGCGAAGTCATGTGGCGTGCACGAAGCGGCATTTCCGCGCTGCTGCGATTTATCT 1533
 DB 61 TCGACGAATCTGTGACAGCGCTCCACGCGCATGGCATGAATGCTGCTGACGCTGCTG 120

QY 1534 TCACACACCTCCAAACGACAGATGGGCGCAACGCTGGCGCGCGGCGGACCGCTTT 1593
 DB 121 CGAAGCACACCTCGAATCATGACGAGTATGTTCCAGCGGCGCATGACCGCGGAACCGGC 180

QY 1594 TCGAC 1598
 DB 181 TCGCC 185

RESULT 8 CB997861/c

LOCUS CB997861 763 bp mRNA linear EST 01-MAY-2003
 DEFINITION AGENCOURT_13906064 NIH_MGC_148 Homo sapiens cDNA clone
 IMAGE:30349222 5', mRNA sequence.

ACCESSION CB997861.1 GI:30292381

VERSION CB997861.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 763)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapps-remail.nih.gov

Tissue Procurement: Dr. Stefan Hansson

cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help and advice from Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: NDAM393 row: c column: 23
 High quality sequence stop: 584.

FEATURES source

1. 763
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30349222"
 /tissue type="pre-eclamptic placenta"
 /lab_host="DH10B Toba"
 /clone_lib="NIH_MGC_148"
 /note="Organ: placenta; Vector: pBluescriptR, Site 1:
 all-XhoI; Site 2: BamH; Library is oligo-dT primed and directionally cloned using primer
 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 180 a 190 c 244 g 149 t

BASE COUNT ORIGIN

Query Match 1.9%; Score 54.6; DB 14; Length 763;
 Best Local Similarity 49.6%; Pred. No. 0.00085;
 Matches 172; Conservative 0; Mismatches 169; Indels 6; Gaps 1;

QY 1308 GCGCATTTGAAGGGCTTGAAAGATAAAATTCCTATTTCAGAGCTTGGTTGACTTAT 1367
 DB 443 GGCGACTTTAAGGGCTGACGGGAACTCGATTACCTACATGGCTCGGGTGGATTGC 384

QY 1368 CTGCACCTGATGCCCTGTTTAAATGCCCTGAAGCAAAAGGACGGGCTATCGGTC 1427
 DB 383 CTATGGTTACCAACCTTCTATGACTCTCT-----CTCCATGACGGAGTTACGACATT 330

QY 1428 AGCAGTACCGCGATGTCAATCCGGCACTGGGCACATAGCGCACTTGGCGAAGTCATT 1487
 DB 329 CTGACTATCTGCTGGATCCGTGAGGAGCTGGGACCATGAGGACTTTAAGTTTCTT 270

QY 1488 GCTGCGCTGCGACGAAGCGGCAATTCGCGCTGCTGCGATTTATCTTCAACACACCTCC 1547
 DB 269 GACGCGACCCATGACCGTGGCTTCGCGTCTCATCTGACTTCGTCATGAACCAACCTCG 210

QY 1548 AACGAACACGATGGCGCAACGCTGGCGCGCGGACCGCTTTTCGACAAATTTCTAC 1607
 DB 209 GATTCCCATCCGTGGTTCTCAGTCTCAGCGGCCGATCCGATGGTCTTACGGTAACCTAC 150

QY 1608 TATATTTTCCCGACCGCGGATGCCGACCAATACGACCGCACCT 1654
 DB 149 TACGTATGTCGGACACTGACGAGGCTTACTCCGATGCTCGTATCAT 103

RESULT 9 CB996585/c

LOCUS CB996585 845 bp mRNA linear EST 01-MAY-2003
 DEFINITION AGENCOURT_13511438 NIH_MGC_148 Homo sapiens cDNA clone
 IMAGE:3030889 5', mRNA sequence.

ACCESSION CB996585

VERSION CB996585.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 845)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgpbbs@mail.nih.gov
Tissue Procurement: Dr. Stefan Hansson
cDNA Library Preparation: Michael J. Brownstein (NHEKRI) with help
and advice from Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM345 row: h column: 02
High quality sequence stop: 610.

FEATURES source
1. 845
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:30330889"
/tissue_type="pre-eclamptic placenta"
/lab_host="DH10B Tona"
/clone_lib="NIH MGC 148"
/note="Organ: Placenta; Vector: pBluescriptR; Site 1:
all-XhoI; Site 2: BamI; Library is oligo-dT primed and
directionally cloned using primer
5'-TTTTTTTTTTTNN-3', size-selected for average insert
size 2.3 kb and normalized to 10^5. This is a primary
library enriched for full-length clones and constructed
using the Cap-trapper method (Carninci, in preparation).
Library constructed by M. Brownstein (NIMH/NHGRI,
National Institutes of Health). Note: this is a NIH_MGC
Library."

BASE COUNT 200 a 209 c 265 g 171 t

Query Match 1.9%; Score 54.6; DB 14; Length 845;
Best Local Similarity 49.6%; Pred. No. 0.00089;
Matches 172; Conservative 0; Mismatches 169; Indels 6; Gaps 1;

QY 1308 GCGATTGAGGGCTTGAAGATAAAATTCCTTTTCAAGAGCTTGGTTGACTTAT 1367
DB 443 GCGACTTTAAAGGGCTGACGGGAAACTCGATTACTACTAAATGCTCGGGGTGATTCG 384

QY 1368 CTGCACCTGATGCGGTGTTTAATCCCTGAGGCAAAAGACGCGGCTATGGGTC 1427
DB 383 CTATGTTACCACTTCTATGACTCTCTT-----CTCATGACGAGGTACGACATT 330

QY 1428 AGCAGCTACCGGATGTCATTCGGCACTGGGCACTAGGCGACTTGGCGAAGTCATT 1487
DB 329 CGTGACTATCGCTGGATCCGGAGGAGCTGGGCACTTGAAGGTTTTCCT 270

QY 1488 GCTGGCTGACGAAGCGGCATTTCCGCGTGTGCTGATTTATCTTACACACCTCC 1547
DB 269 GACGAGCCCATGACCGTGGCTTCCGCTCATCATGCTGCTGATGACCACTCG 210

QY 1548 AACGACACGAATGGGGCAACGCTGCGCGCGCGGACCGCTTTTTCGACATTTCTAC 1607
DB 209 GATTCCATCCGTTGTTCCAGTCTCTACGCGCGGATCGGATGCTCTTACGGTAACTAC 150

QY 1608 TATATTTTCCCGACCGCGGATGCTCCGACCAATACGACCGCACCT 1654
DB 149 TAGCTATGGTGGGACACTGACGAGGCGCTACTCCGATGCTCTGATCAT 103

RESULT 10
CNS08PQT
LOCUS
DEFINITION
Single read from an extremity of a full-length cDNA clone made from
Anopheles gambiae total adult females. 5-PRIME end of clone
PK0AAA34DD06 of strain 6-9 of Anopheles gambiae (African malaria
mosquito).

ACCESSION BX022785
VERSION BX022785.1 GI:27572005
KEYWORDS HTC.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 912)
Genoscope.
TITLE Direct Submission
AUTHORS Genoscope - Centre National de Sequencage
JOURNAL Submitted (06-JAN-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr

FEATURES source
1. 912
Location/Qualifiers
/organism="Anopheles gambiae"
/mol_type="mrna"
/strain="6-9"
/db_xref="taxon:7165"
/clone="PK0AAA34DD06"
/plasmid="pME18S-PL"
/note="end : 5-PRIME"

BASE COUNT 214 a 231 c 290 g 177 t

ORIGIN
Query Match 1.8%; Score 53; DB 11; Length 912;
Best Local Similarity 53.3%; Pred. No. 0.0026;
Matches 138; Conservative 0; Mismatches 115; Indels 6; Gaps 1;

QY 1304 TCGCGGCGATTGAGGGCTTGAAGATAAAATTCCTTTTCAAGAGCTTGGTTGAC 1363
DB 144 TGTGGCGCATCTGACCGGCATTACGGAGAGCTGGCAATTTGGCCGATCTTGGTGTGAC 203

QY 1364 TTACTGCACTGATGCGCTGTTTAATGCTTGAAGCAAAAGCGGCGCTATGC 1423
DB 204 TGGGTGTGGCTGAGCCCGGTGTTTAATCAACCATGCGG-----GACTTGGGTACGA 257

QY 1424 GTCAGCAGCTACCGCATGTCATTCGGCACTGGGCACTGGGCAATAGCGACTTGGCGGAAGT 1483
DB 258 TATTTGGGACTTCCGCGATGGATCCCAATCTTTGGCACTATGGCCGATTTGGACCGTAT 317

QY 1484 CATCTCGCTGTCAGGAAGCGGCAATTCGCGCTGCTCGATTTTATCTTCAACACAC 1543
DB 318 GGTGCAAGGCAAGACGCTCGGTATTAGTGATTTGGACTTTGTGCGGACCATAC 377

QY 1544 CTCCAACGAACACGAATCG 1562
DB 378 GAGCAGCAGCAGCATGG 396

RESULT 11
BX381961/c
LOCUS
DEFINITION
BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0DI072YF05 3-PRIME, mRNA sequence.

ACCESSION BX381961
VERSION BX381961.1 GI:30453007
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Unpublished
Full-length cDNA libraries and normalization
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: secref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0D1072C03NP1.

FEATURES

source
1. .1201
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1072YF05"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 95 a 191 c 115 g 55 t 745 others

ORIGIN

Query Match 1.7%; Score 50.4; DB 13; Length 1201;
Best Local Similarity 7.7%; Pred. No. 0.016;
Matches 70; Conservative 286; Mismatches 556; Indels 0; Gaps 0;

481 GGTGCGGCGGAGTGTGCGTGTGGCGCAATCTTCGCGATGCTGCCGCGCATTCCTCCAA 540

1201 GKGCCGCGSGKGGCGCCCGCCGCCCMVMVMVMVMVMVMVMVMVMVMVMVMVMVMVM 1142

541 AAATCGGCGCGGAACTCGACGATATCAAGAGGAGGAGTGTGATTTGGTCAAGTACGGCG 600

1141 MVMVMGMBVMVGVGKGGKGGKGMVMVMVMVMVMVMVMVMVMVMVMVMVMVMVMVM 1082

601 GAAAGTACGGCGCATCGCGGGCGACAAGCGCACGGCGATTTGGTGGTTCGCTTCGG 660

1081 MSKGVVVRGMBMNMKKGKMSKVMVMVMVMVMVMVMVMVMVMVMVMVMVMVMVM 1022

661 ATAAACGGTGTGTTGACGGTCAAGTGTGTGATGCGCGGTGTTGCGCGCTGAAAGGAC 720

1021 KXEMKTGKKKKKKKVM 962

721 GGTAAAGACGCAATATAGCAAGGACAGGCAATGTTTCAGACGCGATTTCTGTGCGCG 780

961 GGGKMKMKKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMKGMKGM 902

781 CGGCTTGATATGAATCAAGCAGCATCGCGCATATCGGAATCGACATTTGGCAACAGCCTG 840

901 MKKGNMVM 842

841 TCTTTTCTAGTCAGTCGCGAGTCTTTGCGAGTATGATTCAGCAGCAGCGCCTACAGGCAT 900

901 TGCAGGATACGGCGCAGCAGCGCGCTCGGAACTTCAGATCGGAGCAGCATCATGT 960

781 NNN 722

961 TGACCCCGCAGCAGCAAGTCGGTGTGATTTTACAGTACCTTCAAAACACGCACTTTGGACA 1020

721 MTNN 662

1021 TCTACAGCGCCGACAGCGCGCGCGCATCGAAAATCGAAGACTCGCGCGAGTTTTCG 1080

661 NNN 602

1081 GCGCGATGATAGCATCTTCCCAAACTGATGAACGAACTCGACAGCGTGTACGGCAACA 1140

601 NNN 542

1141 AGAAGCCCTGCTGCTGAGAAATGCTGCTGCGGAGGATGCGGCAAGACTATTCC 1200

541 MNNKTKKNN 482

1201 AAGCAACTCATCTCTTAAAGATATCGATATCGCGCGGAAACACACCGCGATTTGATTT 1260

481 KNN 422

QY 1261 TGTCCAAACAACAAGTCGGCGGCTGTAGTGTGATTGTTTTCGCGCGATTGAAGG 1320
DB 421 KGGMMNN 362
QY 1321 GCTTGAAGAATAAAATCTCTTATTTTCAAGAGCTGTGTTGACTTATCTGCACTGATGC 1380
DB 361 NTTTTMMNTNCCNNNTTCANTNTKTTTNNNTTTCNNNNSTCCANATAGNNGGCGGSSASS 302
QY 1381 CGCTGTTTAAAT 1392
DB 301 CGCTGTAGANK 290

RESULT 12

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd,

Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, estfruitfly@berkeley.edu

hit genomic sequence DS04362; hit genomic sequence DS06189

Plate: 63 row: B column: 6

High quality sequence stop: 515.

Location/Qualifiers

1. .659

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="GH06318"

/sex="male and female"

/dev stage="adult"

/lab host="DH5 - alpha"

/clone_lib="GH Drosophila melanogaster head pot2"

/note="Organ: head; Vector: POT2; Site: 1: EcoRI; Site: 2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

BASE COUNT 148 a 160 c 191 g 159 t

ORIGIN

Query Match

Best Local Similarity

Matches 255; Conservative

1.7%; Score 49.8; DB 9; Length 659;

47.4%; Pred. No. 0.017;

Mismatches 266; Indels 15; Gaps 3;

QY 1307 CGCGATTGGAAGGCTTGAAGATAAAATCTTATTTTCAAGAGCTTGGTTGACTTA 1366

DB 128 CGCGATTGGAAGGATCACTCCAGCTCGCTATCTGGCGGACATGGCATCACGC 187

QY 1367 TCTGCACTGATGCGCTGTTTAAATGCTTGAAGGAAAGCGGCGGCTATCGGT 1426

DB 188 CACCTGGTGTAGTCCCATTTTTCAGTCGCC-----CATGATTGACTTTGGCTATGAT 241

QY 1427 CAGCAGCTACCGCGATGTCATTCGCGACTCGGCAATAGCGCACTTGCAGAGTCA 1486

DB 242 ATCGATTACAGGCAATCCAGCCGAGTATGCGACCATGCGAGGATTTTCGAGGAGCTCAT 301

Qy 1487 TGTGCGCTGCACGAAGCGGCATTTCCGCGTCTGCTGATTTTATCTTCAACACACCTC 1546
 Db 302 CGACACGGCTTTCCAGCTGGGCACTCAAGTGTGTTCTGCACTTTGTCGCCGATCACGCTC 361
 Qy 1547 CAACGAACACGAATGGGCGCAACGCTGCGCGCGCGGCGACCCGCTTTTTCGACATTTCTA 1606
 Db 362 GGATCAGCATGAGTGGTTCAGAAAGTCTGGGCAAGGAGCGCGGTTTACGAGGATTT--- 418
 Qy 1607 CTATATTTTCCCGACCGCGGATGCGCGACCAATACGACCGCACCCCTCGCGGAAATCTT 1666
 Db 419 CTATGTTGGACATGCACTGCTCGAGGAGATGGCACTCGGGTGCACCACTG 478
 Qy 1667 CCGCAGCAGCAGCCGCGCGCTTCTCGCACTCGAAGACGAGCGTGGTGTGGAGAC 1726
 Db 479 GCAATCGGTGTTACGATCCGCTTGGGATGCGACGAAGGTCTGAGCAGTATTACCT 538
 Qy 1727 CTTCAATCTTCTC-----CAATCGGACTTCAATTACAGCAACCGGCTATTCCGCGC 1780
 Db 539 GCACCACTTACCAGGAGACCGGACTTGAATATCGCAATCCCAAGTGGTTCAGGC 598
 Qy 1781 AATGCGGCGGAAATGCTTCTTCTGCACTTGGGCGTTGACATCTCGGTATGGAT 1838
 Db 599 CATGGATGATGTGTTGCTTTCTGCTCAACAGGCGGTACTGCTTTTCGATCGAT 656

RESULT 13
 BM587500
 LOCUS 703 bp mRNA linear EST 25-FEB-2002
 DEFINITION 17000687321295 A.Gam.ad.cdna.blood1 Anopheles gambiae cDNA clone
 19600449698387 5', mRNA sequence.

ACCESSION BM587500
 VERSION BM587500.1 GI:188893361
 KEYWORDS EST.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles
 1 (bases 1 to 703)
 REFERENCE Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab
 R., Collins, F.H., Venter, J.C. and Hoffman, S.D.
 Celera Anopheles gambiae EST project
 Unpublished
 CONTACT: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltR@celera.com
 Plate: NU01004P70 row: F column: 05
 Seq primer: M13 Reverse.

FEATURES
 source
 1..703
 /organism="Anopheles gambiae"
 /mol_type="mRNA"
 /strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449698387"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /clone_lib="A.Gam.ad.cdna.blood1"
 /note="vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3' clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"
 BASE COUNT 155 a 178 c 228 g 142 t

Query Match 1.7%; Score 49.8; DB 12; Length 703;
 Best Local Similarity 52.5%; Pred. No. 0.018;

Matches 136; Conservative 0; Mismatches 117; Indels 6; Gaps 1;
 Qy 1304 TCCCGCGATTTGAAGGGCTTGAAGATAAATCTCTTATTTTCAAGAGCTTGGTTGAC 1363
 Db 147 TGTGGCGGATCTGACCGGCAATTACGGAGAGCTGGAACATTTTGGCCGATCTTGGTGTGAC 206
 Qy 1364 TTATCTGCACCTGATGCCGCTGTTTAAATGCCCTGAAGCAAAAGCGACGGCGCTATGC 1423
 Db 207 TGGGGTGTGGCTGAGCCCGGTGTTTAAATCAACCGATGGCG-----GACTTTGGGTACGA 260
 Qy 1424 GGTACAGCACTACCGGATGTCATCCGCACTGGGCAATAGGCGACTTGGCGAAGT 1483
 Db 261 TATTTGCGACTTCCCGGATGTGATCCCAATCTTTTGGCACTATGGCCGATTTGACCGTAT 320
 Qy 1484 CATTTGCTGGCTGCGACGAAGCCGCAATTCGCCCTGTCGATTTTATCTTCAACACAC 1543
 Db 321 GTTGCAGAGGCAAGACGCTCGTATTAAAGGTGATTTCTGGACTTTTGTCCGAATCATAC 380
 Qy 1544 CTCCACAGCAACGATGG 1562
 Db 381 GAGTGACGAGCAGATGG 399

RESULT 14
 BH770895/c
 LOCUS 1981 bp DNA linear GSS 01-MAY-2002
 DEFINITION LLMcag529 MG1363 Random Sequence Tag Library Lactococcus lactis
 subsp. cremoris genomic, genomic survey sequence.

ACCESSION BH770895
 VERSION BH770895.1 GI:20373852
 KEYWORDS GSS.
 SOURCE Lactococcus lactis subsp. cremoris
 ORGANISM Lactococcus lactis subsp. cremoris
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Lactococcus
 1 (bases 1 to 1981)
 REFERENCE Bolotin, A., Ehrlich, S.D. and Sorokin, A.
 A. Studies of genomes of dairy bacteria Lactococcus lactis
 Sci. Aliments, (2002) In press
 CONTACT: Sorokin, A.
 Genetique Microbiome
 INRA
 CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
 Tel: 33 1 34 65 25 15
 Fax: 33 1 34 65 25 21
 Email: sorokine@jouy.inra.fr
 best homologue in strain IL1403 is amyV (90%)
 Class: shotgun
 High quality sequence start: 30
 High quality sequence stop: 1953.
 Location/Qualifiers
 1..1981
 /organism="Lactococcus lactis subsp. cremoris"
 /mol_type="genomic DNA"
 /strain="MG1363"
 /db_xref="taxon:1359"
 /clone_lib="MG1363 Random Sequence Tag Library"
 /note="vector: pSGM2; Site 1: SmaI; Library of chromosomal fragments of L.lactis strain MG1363 was prepared by partial AluI digestion or by sonication."

BASE COUNT 574 a 429 c 335 g 643 t
 ORIGIN

Query Match 1.7%; Score 49.8; DB 28; Length 1981;
 Best Local Similarity 54.7%; Pred. No. 0.03;
 Matches 99; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 1406 AAGCGAGCGCGGTATGCGGTACGAGCTACCGGATGTCAATCCCGCACTGGGCACAAT 1465
 Db 1634 AAGCTATCATGCTATGATGTACGATACGAGAAATTAATCTTAATTTGGCAAT 1575
 Qy 1466 AAGCGACTTGGCGAAGTCAATTCCTCGCTGCGACGAAGCGGCAATTCGCGCTCGA 1525

Db 1574 GGCTGATTTTGAATACTGATTCCTCAGGCTATAAAACAGGAATTCCTGCGTGAATTTTGA 1515
QY 1526 TTATTACTTCAACACACCTCCACACACACGATGGCCACACGCTGCGCCCGCGGA 1585
Db 1514 CATCGCGTTTATCCACACGCGACTGACAAATGCTGGTTTCAAAAAGCAGTACTAGCTGCGGA 1455
QY 1586 C 1586
Db 1454 C 1454

RESULT 15

AI386572
LOCUS
DEFINITION
GH16695.5prime GH Drosophila melanogaster head pOT2 Drosophila
melanogaster cDNA clone GH16695 5prime similar to V00204: Lvph
PBgn0002570 PID:g7777 SWISS-PROT:P07190, mRNA sequence.
AI386572
AI386572.1 GI:4200583
EST.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera, Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 551)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic sequence AC005430: hit genomic sequence AC005451
Plate: 166 row: H column: 11
High quality sequence stop: 544.
Location/Qualifiers
1. .551
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GH16695"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/clone_lib="GH Drosophila melanogaster head pOT2"
/notes="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT 139 a 152 c 150 g 110 t

FEATURES
source

Query Match 1.7%; Score 49.2; DB 9; Length 551;
Best Local Similarity 50.3%; Pred. No. 0.024;
Matches 151; Conservative 0; Mismatches 143; Indels 6; Gaps 1;
QY 1308 GCGGATTTGAAGGCTTGAAGATAAAATCTTATTTTCAGAGCTTGGTTTGACTTAT 1367
Db 128 GCGCACTTGAACGGGTCTACTGAAAGCTGCAGTACTCTGAAGACATCGGCTTCACGGGC 187
QY 1368 CTGCACCTCATCGCGTGTTTAAATGCCCTGAAGGCAAAAGCGGGGCTATCGGTC 1427
Db 188 ACATGGCTCTGCCCATATTCAATCGCC-----CATGGTCGCTTTGGTTACACATA 241
QY 1428 AGCAGTACCGGATGTCAATCGGCACCTGGGACAAATAGGCGACTTGGCGAAGTCAAT 1487
Db 242 TCGGACTTCTACAGATCCATCCGAATATGAAACCATGGAGACTTTGAGCGAATGATC 301
QY 1488 GTTGGCTGCGAGACCGGCAATTCGGCGCTGCTGATTTATCTTCAACCGACCTCC 1547

Db 302 GCCAGGCCACAGGAGTGGGCATTAAATCATCTCTGGACTTCGTACCAACCACTCAAGT 361
QY 1548 AACGAACACGAATGGCGGCAACGCTGGCGCGCGGAGCCCGCTTTTCGACAAATTTCTAC 1607
Db 362 ACCGAAAACGAATGGTTTCAACAAGTCTGTGGACAGTGCACCGCTCTACAAAGACTTCTAC 421

Search completed: November 8, 2003, 17:31:23
Job time : 5420 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2003, 17:35:13 ; Search time 5434 Seconds
(without alignments)

4788.098 Million cell updates/sec

Title: US-09-843-007A-2

Perfect score: 3408

Sequence: 1 MLTPQQVGLIQLKTRIL.....VSLNQDLTLQPYQVMWLEIA 636

Scoring table:

BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DRV=xlp
-Q/cgn2_1/USPTO.spool_p/US09843007/runat_07112003_170420_25346/app_query.fasta_1.775
-DB=GenEmbl -QFWT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09843007@cgn_1_1_4958@runat_07112003_170420_25346 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -JONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: gb.da.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sv.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3408	100.0	1939	6	AXI07013	Sequence
2	3408	100.0	2115	1	NPO011781	AXI017013 Sequence
3	3408	100.0	6878	6	AXI07015	Sequence
4	3360	98.6	1910	6	AXI07014	Sequence
5	3360	98.6	6851	6	AXI07016	Sequence
6	3338	97.9	1967	1	AY099334	Neisseria
7	3333	97.8	1967	1	AY099335	Neisseria
8	3141.5	92.2	2883	6	A47369	Sequence 1
9	3141.5	92.2	2883	6	A47785	Sequence 1
10	1319.5	38.7	10719	1	AE001346	AE001946 Peinococc
11	1126.5	33.1	10261	1	AE011997	AE011997 Xanthomon
12	1116.5	32.8	11605	1	AE012454	AE012454 Xanthomon
13	915.5	26.9	10822	1	AE005791	AE005791 Caulobact
14	550	16.1	304517	1	AE016789	AE016789 Pseudomon
15	548.5	16.1	189050	1	AL646077	AL646077 Ralstonia
16	543.5	15.9	17863	1	AE004642	AE004642 Pseudomon
17	529.5	15.5	1794	6	AXI22611	Sequence
18	529.5	15.5	1794	6	BD164728	BD164728 Nevel pol
19	529.5	15.5	1981	6	AX469814	AX469814 Sequence
20	529.5	15.5	9184	1	TSO001206	AP001206 Streptomy
21	529.5	15.5	281450	1	AP005032	AP005032 Streptomy
22	529.5	15.5	303550	1	SC0939131	AL939131 Streptomy
23	529.5	15.5	326150	1	AP005281	AP005281 Corynebac
24	529.5	15.5	349980	6	AXI27150	Sequence
25	529.5	15.5	349980	6	AXI27151	Sequence
26	529	15.5	11866	1	AE012109	AE012109 Xanthomon
27	526.5	15.4	306650	1	AP005221	AP005221 Corynebac
28	525.5	15.4	10134	1	AE012958	AE012958 Chlorobiu
29	523.5	15.4	12548	1	AE011639	AE011639 Xanthomon
30	522	15.3	298600	1	AP005959	AP005959 Bradyrhiz
31	521.5	15.3	309904	1	AE016865	AE016865 Pseudomon
32	520.5	15.3	10890	1	AE002040	AE002040 Deinococc
33	517.5	15.2	10718	1	AE010836	AE010836 Methanosa
34	516	15.1	9589	1	SCO093123	AL939123 Streptomy
35	515.5	15.1	300100	1	SC0939123	AL939123 Streptomy
36	515	15.1	4121	1	AF135796	AF135796 Thermus c
37	513	15.1	2895	6	BD180010	BD180010 Highly th
38	511.5	15.0	325483	1	AP005050	AP005050 Streptomy
39	509.5	15.0	11700	1	MTC14188	Z96071 Mycobacteri
40	509.5	15.0	14029	1	AE006925	AE006925 Mycobacte
41	509.5	15.0	343050	1	BX248334	BX248334 Mycobacte
42	504	14.8	1704	6	AR012159	AR012159 Sequence
43	504	14.8	1704	6	AR012160	AR012160 Sequence
44	504	14.8	1704	6	E10495	E10495 gDNA encodi
45	504	14.8	2304	1	D78198	D78198 Pimelobacte

ALIGNMENTS

RESULT 1

AX107013	AX107013	Sequence 4 from Patent WO0125449.	1939 bp	DNA	linear	PAT 30-APR-2001
LOCUS	AX107013	AX107013				
DEFINITION	AX107013	AX107013				
ACCESSION	AX107013	AX107013				
VERSION	AX107013.1	GI:13922547				
KEYWORDS		synthetic construct				
SOURCE		synthetic construct				
ORGANISM		artificial sequences.				
REFERENCE						
AUTHORS		Bengs, H., Polakowski, T., Held, A. and Gallert, K.C.				
TITLE		Novel immobilizable amylase, the use thereof and method for				
		producing poly(1,4-alpha-glucan)				
JOURNAL		Patent: WO 0125449-A 4 12-APR-2001;				
		Axiva GmbH (DE)				
FEATURES						
source		Location/Qualifiers				
		1..1939				
		/organism="synthetic construct"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:32630"				
		/note="PCR-Amplikon, kodierend fur amylase (Amsu)"				
BASE COUNT	453 a	605 c	482 g	395 t	4 others	
ORIGIN						
Alignment Scores:						
Pred. No.:	5,65e-287	Length:	1939			
Score:	3408.00	Matches:	636			
Percent Similarity:	100.00%	Conservative:	0			
Best Local Similarity:	100.00%	Mismatches:	0			
Query Match:	100.00%	Indels:	0			
DB:	5	Gaps:	0			
US-09-843-007A-2 (1-636) x AX107013 (1-1939)						
Qy	1	MetLeuThrProThrGlnValGlyLeuLeuGlnTyrLeuLysThrArgGlyLeu	20			
Db	12	ATGTTGACCCCGCCAGAGTGGTTCATTTACAGTACCTCAAAACACGACATCTG	71			
Qy	21	AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspThrArgGlnPhe	40			
Db	72	GACATCTACAGCGCCGACAGCGCGCGGATCGGATCGGAAATCCGAAAGACTGCGCGCAGTTT	131			
Qy	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60			
Db	132	TCGCGCGCGATGGATAGCATTTCCCAACATGATGACGACCTCGACAGCGTGTACGGC	191			
Qy	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaThrGlnSerTyr	80			
Db	192	AACAACGAAGCCCTGCTGCTATCTGGAATGCTGCTGGCGCAGCGCATGCGCAAGCTAT	251			
Qy	81	SerGlnArgSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspThr	100			
Db	252	TCCCAACGCACTCATCTTAAAGATATCGATATCGCGCGGAAACAAACCCCGATGG	311			
Qy	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120			
Db	312	ATTTTGTCCAAACAAACAAAGTCGGCGGGGTGTGTACGTTGATTTGTTTGGCGGATTTG	371			
Qy	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu	140			
Db	372	AAGCGCTTGAAGATAAAATTCCTATTTCAGAGCTTGGTTTACCTTATCTGCACCTG	431			
Qy	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr	160			
Db	432	ATGCGCTGTTTAATGCCCTGAGGCAAGACGCGCGGCTATGCGGTGAGCAGCTAC	491			
Qy	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu	180			
Db	492	CGCATGTTCATCCGCGACATGGGACCAATAGCGGACTTGGCGAAGTCATTGCTGGCGCTG	551			
Qy	181	HisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHis	200			
Db	552	CACGAGCCGCAATTCGCGCTGTCGATTTTATCTCAACACACCTCCACGACAC	611			

Qy	201	GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220			
Db	612	GAATGGGCGCAACGCTGCGCGCGCGCGCGCGCTTTTCGACAAATTCCTACTATATTC	671			
Qy	221	ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240			
Db	672	CCCGACCGCGGATGCGGACCAATACGACCGCACCTGCGGAAATCTTCGCCGACAG	731			
Qy	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTyrThrThrPheAsnSer	260			
Db	732	CACCGCGCGGCTTTCGCACTGGAGCGACGCTGGTGTGGACGACCTTCATATTC	791			
Qy	261	PheGlnTrpAspLeuAsnTyrSerAsnProThrValPheArgAlaMetAlaGlyGluMet	280			
Db	792	TTCCAATGGGACTTGAATATACAGCAACCGTGGTATTCCGCGCAATGCGCGCGAAATG	851			
Qy	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr	300			
Db	852	CTGTTCTTCGCAACTTGGCGCTTGACATCTTCGCTGATGATGCGGTGCTTATTTGG	911			
Qy	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320			
Db	912	AAACAAATGGGACCAAGCTGCGAAACCTGCGCGAGGCGACGCCCTCATCCGCGCTTC	971			
Qy	321	AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340			
Db	972	AATGCGCTTATCGTATTCGCGCGCGCGCGCTTCTTCAATCCGAAACGCTTCCTCAC	1031			
Qy	341	ProAspGlnValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360			
Db	1032	CCCGACCAAGTCGTCCATACATCGGCGAGGACGAATGCGCAATCGGTTCACACCCCTG	1091			
Qy	361	GlnMetAlaLeuLeuThrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla	380			
Db	1092	CAATGGCATTTGTTGGAAACCTTCGACGCGCGGAAAGTCAACTGCTTCATCAGGGG	1151			
Qy	381	LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis	400			
Db	1152	CTGACCTACCGCAACACTGCGCGAGCATACGCTGCTGCTGCTGCTGCTGCTGCTGCT	1211			
Qy	401	AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr	420			
Db	1212	GACGACATCGGCTGGACGTTCCGATGAGACGCGGCAATCTGCGGCAATAAGCGGTAC	1271			
Qy	421	AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg	440			
Db	1272	GACACCGCGCAATTCCTCAACCGCTTCCTGCAACCGCTTCGACGCGCAGCTTCGCTCGT	1331			
Qy	441	GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla	460			
Db	1332	GGCGTACCGTTCCATATCAACCAAGCACAGCGACTGCGGTGTCAGTGTGTACAGCGCG	1391			
Qy	461	AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr	480			
Db	1392	GCATTGCTGCGCTTGGCGAAGCATCCCGCGGTTCGCGCATCAAACTCTTGTATC	1451			
Qy	481	SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500			
Db	1452	AGCATTTGCTTGTAGTACCGCGCTCTGCGCTGATTTACCTAGCGCACCAAGTGGGTACG	1511			
Qy	501	LeuAsnAspAspThrProSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis	520			
Db	1512	CTCAATGACGACGACTGGTCCGAAGACAGCAATAGAGCAGCAGCGCTTGGCGCGAC	1571			
Qy	521	ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540			
Db	1572	CGTCCCGGTACACGACGCTCTGACGCGCAACGATCCGTCGACCGCGCGCGCG	1631			
Qy	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560			
Db	1632	CAAATCTATCAGGCTTGGCGCATATGATTCGCGTCCGCAAGCAATCCGCGCTTCGAC	1691			

QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleileGlyTyrIleArgAsn 580
Db 1692 GGCGGAGGCTGGTTACATTCACCAACCAACAGCACATCATCGGTACATCCGCAAC 1751
QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGlyThrProGlnThrValThrAlaHisThr 600
Db 1752 AATGGCGCTTTGGCATTCGGTAACTTCAGCGAATATCCGCAACCGTTACCGCGCATACC 1811
QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
Db 1812 CTGCGAGCATGCCCTTCAGCGCGCAGACCTCATCGTGGCGCAAACTGTACGCTGAT 1871
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
Db 1872 CAGGATTTGACGCTTCAGCGCTTATCAGGTCAATGCGCTCGAATCGCC 1919
RESULT 2
NPO011781
LOCUS 2115 bp DNA linear BCT 28-JAN-1999
DEFINITION Neisseria polysaccharaea amylosucrase gene.
ACCESSION AJ011781
VERSION AJ011781.1 GI:4049491
KEYWORDS amylosucrase; amylosucrase gene.
SOURCE Neisseria polysaccharaea
ORGANISM Neisseria polysaccharaea
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE 1
AUTHORS De Montalk G.P., Remaud-Simeon M., Willemot R.M., Planchot V. and
Monsan P.
TITLE Sequence analysis of the gene encoding amylosucrase from Neisseria
polysaccharaea and characterization of the recombinant enzyme
J. Bacteriol. 181 (2), 375-381 (1999)
MEDLINE 99102197
PUBMED 9882648
REFERENCE 2 (bases 1 to 2115)
AUTHORS Potocki de Montalk G.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-1998) Potocki de Montalk G., U.M.R. C.N.R.S.
5504, L.A. I.N.R.A. D.G.B.A., I.N.S.A., Centre de bioingenierie
Gilbert Durand, Complexe scientifique de Rangueil, Toulouse cedex,
31077, FRANCE
FEATURES
source
1..2115
/organism="Neisseria polysaccharaea"
/mol_type="genomic DNA"
/strain="ATCC 43768"
/db_xref="taxon:489"
136..2046
/gene="amylosucrase"
<136..2046
/gene="amylosucrase"
/EC_number="2.4.1.4"
/codon_start=1
/transl_table=11
/product="amylosucrase"
/protein_id="CAA09772.1"
/db_xref="GI:4107260"
translation="MLTPQQVGLILQVLTILDIYTPQAPAGIEKSDMRQSRRM
DTHFKMLNLSVGVNREALLPMLELLAQWQSYQSNLSLKDIDIAENRDPWIL
SNKQGVGVCTVDFAGDLKGLDKI FYQELGLTILHLPLFCPEGSDGQYVSSY
RDVNFALGTILRLVIALHEAGISAVVDYFNFHNSHEHWAQSCAAGDPLFDNFY
IFPDRMPQYDRLREIFPDQPGQPSQLEDGRVWTFNSFQNDLNYSNPWFPM
ACEMFLQVGLIDILMDAVFIWKQGTSCENLPQAHALIAFNVAIVPAVFFK
SEAIYHPQVQVIGQECQIGYNPLQWALLNTRVNLHQAALYRHLNPHSTA
WNYVSRDDIGWTADDAAYLIGISVDHROPFNRFVNRFDGSAFSGVFPYNPST
GCRVSGTAAUUGLAQDDPHAVDRIKLLYSIALSTGGPLIYLIGDEVTLNDDWSQ
DSNKSDSWAHRPRYNEALYQADNPSTAAGQIYQGLNHLIAVQSNRPFQDGRKLT
FNTNKKHIGIYRNALLAFGNFSEYPTQVTTAHTLQAMPFFKAHDLIGGKTLSLQNDLT
LQPYQVMLEIA"
493 a 529 g 432 t
BASE COUNT 493 a 529 g 432 t
ORIGIN

Alignment Scores:
Pred. No.: 6,39e-287 Length: 2115
Score: 3408.00 Matches: 636
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-09-843-007A-2 (1-636) x NPO011781 (1-2115)
QY 1 MetLeuThrProThrGlnGlnValGlyLeuLeuGlnTyrLeuLysThrArgIleLeu 20
Db 136 ATGTTGACCCCAACGACCAAGTGGTTTGATTTTACAGTACCTCAAAACAGCATCTTG 195
QY 21 AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe 40
Db 196 CACATCTACACGCCGCAACAGCGCGCCGACATCGAAAAATCCGAAGACTGCGCGCAGTTT 255
QY 41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
Db 256 TCGCGCCGCATGGATACGCAATTTCCCHACTGATGAACGAACTCGACAGGCTGTACGCG 315
QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr 80
Db 316 AACCAACGAAGCCCTGCTGCTATGCTGCAATGCTGCTGCGCGCAGGATGCGAAAGCTAT 375
QY 81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrp 100
Db 376 TCCCAACGCAACTCATCTTAAAGATATCCATATCGCGCGCAAAACAAACCCGATTTGG 435
QY 101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
Db 436 ATTTGTCCCAACAACAAGTCGCGCGCTGTGCTACGTTGATTTGTTTCCGCGGATTTG 495
QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrIleHisLeu 140
Db 496 AAGGCTTTGAAGATAAAATTCCTTATTTCAAGAGCTTGGTTGACTATCTGCACCTG 555
QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
Db 556 ATGCGCTGTTTAAATGCCCTTGAAGCAACGCGCGCTATGCGTCAGCAGCTAC 615
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
Db 616 CGCATGTGCAATCCGCGCACTGGGCACAATAGCGCACTTTCGCGAAGTCAATTCGCTG 675
QY 181 HisGluAlaGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSerAsnGluHis 200
Db 676 CACGAAGCGCGCATTTTCGCGCTGCTGATTTTATCTTTCAACCAACACCTCCCAACGAC 735
QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
Db 736 GAATGGCGCAACGCTGCGCGCGCGCGCGCGCTTTTCGACAATTTCTACTATATTTC 795
QY 221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
Db 796 CCCGACCGCGGATGCCCGACCAATACGACCGACCCCTTCGCGCAAAATCTCCCGACAC 855
QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsnSer 260
Db 856 CACCGCGCGCTTCTCCCACTGGAAGACGACGCTGGGTGTGGACGACCTCAATTCC 915
QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280
Db 916 TTCCAATGGGACTTGAATTAACGCAACCCCGTGGGTATTTCGCGCAATGCGCGCGGAAATG 975
QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300
Db 976 CTGTTCTTCCTGCAACTTGGCGGTGTGACATCTCTGATGATGATGCGGTTCCTTTATTGG 1035
QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
Db 1036 AACCAATGGGCAACAGCTGCGAAAACTGCGCGAGGCGCACGCGCTCATCCGCGCGTTC 1095

ORGANISM	synthetic construct	artificial sequences.
REFERENCE	1	
AUTHORS	Bengs H., Polakowski T., Heid, A. and Gallert, K.C.	
TITLE	Novel immobilizable amylosucrase, the use thereof and method for producing poly(1,4-alpha-glucan)	
JOURNAL	Patent: WO 0125449-A 6 12-APR-2001;	
FEATURES	Axiva GmbH (DE)	
source	Location/Qualifiers	
	1..6878	
	/organism="synthetic construct"	
	/mol type="genomic DNA"	
	/db xref="taxon:32630"	
	/note="Beschreibung der kunstlichen	
	Sequenz: Expressionsvektor pGEX-4T-1-Amsu"	
BASE COUNT	1673 a 1800 c 1766 g 1639 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	3,36e-286	Length: 6878
Score:	3408.00	Matches: 636
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	6	Gaps: 0
US-09-843-007A-2 (1-636) x AX107015 (1-6878)		
QY	1 MetLeuThrProThrGlnGlnValGlyLeuLeuGlnTyrLeuLysThrArgIleLeu 20	
DB	945 ATGTTGACCCCCACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1004	
QY	21 AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe 40	
DB	1005 GACATCTACACGCCGGAACACGACGACGACGACGACGACGACGACGACGACGACG 1064	
QY	41 SerArgGlnMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60	
DB	1065 TCGCCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1124	
QY	61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr 80	
DB	1125 AACACGAAGCCCT 1184	
QY	81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTyr 100	
DB	1185 TCCACACGCACT 1244	
QY	101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120	
DB	1245 ATTTTGTCCAAACAACACAGTCCGCGCGTGTGCTACGTTCATTTGTTTGGCGGCA 1304	
QY	121 LysGlyLeuLysAspLysIleProTyrPheGlnGlnLeuGlyLeuThrTyrLeuHisLeu 140	
DB	1305 AAGGCTTCAAGAGATAAATTCCTTATTTTCAAGAGCTTGGTTTGTGACTTATCTG 1364	
QY	141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160	
DB	1365 ATGCCCTCTTTTAAATGTCCTGAGGCAAAAGCGACGCGCTATGCGTTCAGCAGCT 1424	
QY	161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu 180	
DB	1425 CGCGATGTCAATCCGCACTGGGCAATAGCGCTTTCGCGCAAGTCAATGCTGCGCGCT 1484	
QY	181 HisGluAlaGlyLysSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200	
DB	1485 CAGAGAGCGGCAATTCGCGCTCTCTGATTTATCTTCAACCCACACCTCCACGAA 1544	
QY	201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220	
DB	1545 GAATGGCGCAACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1604	
QY	221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240	

Db 1605 CCGACCGCGGATGCCGACCAATACGACGACCTCTGGCGAATCTTCCCGACCG 1664
Qy 241 HisproGlycylPheSerGlnLeuGluAspGlyArgTyrValThrThrPheAsnSer 260
Db 1665 CACCGCGCGGCTTCTCGCAACTGGAGACGCGCTGGGTGTGGACGACCTTCAATTC 1724
Qy 261 PheGlnThrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet 280
Db 1725 TTCATATGGGACCTTGAATATCAGCAACCGGTGGGTATTCGCGCAATGGCGGGAATG 1784
Qy 281 LeuPheLeuAlaAsnLeuGlyValAspLeuArgMetAspAlaValAlaPheIleTyr 300
Db 1785 CTGTTCCTTCCCACTTGGCGGTGGACATCTCTGCTATGATGCGGTGCTTATTTGG 1844
Qy 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
Db 1845 AACCAATATGGGACAGCTGGAAACCTGCGGACGCGGACGCTCATCCGCGCGCTTC 1904
Qy 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
Db 1905 AATGCGGTATGCGTATGCGCGCGCGCGGTGTCTTCAATCCGAAGCATCGTCCAC 1964
Qy 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
Db 1965 CCCGACCAAGTCGTCCAAATACATCGGCGAGCAAGTCCAAATCGGTTCACACCCCTG 2024
Qy 361 GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380
Db 2025 CAAATGGCATTTGTTGGACACCTTGTGCGCGCGGAGTCAACCTGCTCCATCAGCG 2084
Qy 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis 400
Db 2085 CTGACCTACCGCCACCACTGCCCGGACATACCGCTGGGTCAACTAGTCCGAGCGCAC 2144
Qy 401 AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420
Db 2145 GACGACATCGCTGGACGCTTGGCGATGAAGACGCGGATATCTGGGCTAAGCGGTAC 2204
Qy 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
Db 2205 GACCAACCGCATCTCTCAACCGCTTCTGCTCAACCGTTTCGACGCGGAGCTTCGCTGT 2264
Qy 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
Db 2265 GCGGTACCGCTTCCAATACACCCAGACAGCGAGTCCGCTGTGTCAGTGGTACAGCGCG 2324
Qy 461 AlaLeuValGlyLeuAlaGlnAspProHisAlaValAspArgIleLysLeuLeuTyr 480
Db 2325 GCAATGGTGGCTTGGCGCAAGACCATCCCAACGCGGTGTACCGCATCAACTCTTGTAC 2384
Qy 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
Db 2385 AGCATTTGCTTGGTACCGCGGTCTGCGCGCTGATTTACCTAGGCGACGAAAGTGGTACG 2444
Qy 501 LeuAsnAspAsnTyrSerGlnAspSerLysSerAspAspSerArgTyrAlaHis 520
Db 2445 CTCATATGACGACGTGTGCGAAGACAGCAATATAGCGGACGACGCGGTGGCGGCAC 2504
Qy 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
Db 2505 CGTCCGCGCTACACGAGCGCTGTACGCGCAACGCAACGATCCGTGCGACGCGCGG 2564
Qy 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
Db 2565 CAAATCTATCAGGCTTGGCGCATATGATGTCGTCGCGCAAGCAATCCGCGCTTCGAC 2624
Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
Db 2625 GCGCGAGGCTGGTATCATTCACACCAACACAGCACATCATCGCTACATCCGCAAC 2684
Qy 581 AsnAlaLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600

Db 2685 AATGCGCTTTGGCATTCGGTAACCTTCAGCGAATATCCCAACCGTTTACCGCGCATACC 2744
Qy 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
Db 2745 CTGCAAGCCATGCCCTTCAGGCGACAGACCTTCATCGTGGCAAACTGTCAAGCTGAAT 2804
Qy 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla 636
Db 2805 CAGGAATTCAGCTTCAGCCCTATCAGGTTCATGTGCTCGAATCGCC 2852
RESULT 4
AX107014 1910 bp DNA linear PAT 30-APR-2001
LOCUS Sequence 5 from Patent WO0125449.
DEFINITION AX107014
ACCESSION AX107014.1 GI:13922548
VERSION
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Bangs,H., Polakowski,T., Heid,A. and Gallert,K.C.
TITLE Novel immobilizable amylosucrase, the use thereof and method for
producing poly(1,4-alpha-glucan)
JOURNAL Patent: WO 0125449-A 5 12-APR-2001;
Axiva GmbH (DE)
FEATURES Location/Qualifiers
source 1..1910
/organism="synthetic construct"
/mol type="genomic DNA"
/db xref="taxon:32630"
/note="PCR-Amplikon, kodierend fur verkurzte amylosucrase (AMSus)"
BASE COUNT 447 a 595 c 474 g 387 t 7 others
ORIGIN
Alignment Scores:
Pred. No.: 8,36e-283 Length: 1910
Score: 3360.00 Matches: 626
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.84% Mismatches: 0
Query Match: 98.59% Indels: 0
DB: Gaps: 6
US-09-843-007A-2 (1-636) x AX107014 (1-1910)
Qy 10 LeuIleLeuGlnTyrLeuLysThrArgIleLeuAspIleTyrThrProGluGlnArgAla 29
Db 10 ATGATTTTACAGTACCTCAAAACACGCGATCTTGACATCTACACGCCCGACACGCGGCC 59
Qy 30 GlyIleGluLysSerGluAspTyrArgGlnPheSerArgArgMetAspThrHisPhePro 49
Db 70 GGCATCCGAAATCCGAGACTGGCGGCACTTTTGGCGCGCATGTGATACGCAATTCGCC 129
Qy 50 LysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsnGluAlaLeuLeuProMetLeu 59
Db 130 AAACGTGATGAACGAACTCGACGCGTGTACGCGCAACAAACGAGCCCTGCTGCCCTATGCTG 189
Qy 70 GluMetLeuLeuAlaGlnAlaTyrGlnSerTyrSerGlnArgAsnSerSerLeuLysAsp 89
Db 190 GAAATGCTGTGGCGCGAGCGCATGCGCAAGTATTTCCCAACGCACTCATCTTAAAGAT 249
Qy 90 IleAspIleAlaArgGluLeuAsnAsnProAspTyrIleLeuSerAsnLysGlnValGlyGly 109
Db 250 ATCGATATCGCGCGGAAACCAACCCCGATGGATTTTGTCCAAACAAACGAGTCGCGGCC 309
Qy 110 ValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyr 129
Db 310 GTGTGCTACGTTGATTTGTTTGGCGCGGATTTGAAGGCTTGAAGATATAATTCCTTAT 369
Qy 130 PheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGluGly 149
Db 370 TTTCAAGAGCTTGGTTTGACTTAICTGCACCTGATGCGCGTGTAAATGCGCTCGAGGC 429

QY 150 LysSerAspGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThr 169
 DB 430 AAAGCGAGCGGGCTATCGGTGACAGCTACCGGAGTGTCAATCCGCACTGGGCACA 489
 QY 170 IleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIleSerAlaVal 189
 DB 490 ATAGCGGACTTGGCGAAGTCATGCTGGCTGCACGAAGCGGCAATTTCCGCGTGTGTC 549
 QY 190 AspPheIlePheAsnHisThrSerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGly 209
 DB 550 GATTTTATCTTCAACCAACCTCCACGACACGAATGGCGCAACGCTGCGCGCGCGC 609
 QY 210 AspProLeuPheAspAsnPheTyrIlePheProAspArgArgMetProAspGlnTyr 229
 DB 610 GACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATAC 669
 QY 230 AspArgThrLeuArgGlnIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGlu 249
 DB 670 GACCGCACCTTGGCGGAATCTTCCCGGACACGACCCGCGCGGCTTCTCGCAACTGGAA 729
 QY 250 AspGlyArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269
 DB 730 GACGAGCGTGGTGTGGAGACCTTCAATCTCTTCCATGGGACTTGAATTACAGCAAC 789
 QY 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
 DB 790 CCGTGGGTATTCGCGCAATGGCGGGCGGAAATGCTGTCTTCCCAACTGGGCGGTGAC 849
 QY 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309
 DB 850 ATCTGCGTATGATGGGTGGCTTGTCTTATTGGAAACAAATGGGACAGCTGCGGAAC 909
 QY 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro 329
 DB 910 CTGCGCAGCGACGCGCTCATCCGCGGCTCAATGCGCTTATGGGTATGTGGCGGCC 969
 QY 330 AlaValPhePheIleSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349
 DB 970 GCGGTGTCTTCAATCCGAAGCGATGCTCCACCCGACCAAGTGTCCATACATCGGG 1029
 QY 350 GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeu 369
 DB 1030 CAGGAGGANTGCCAATCGGTACACCCCTCCAAATGCCATTTGTGGACACCCCT 1089
 QY 370 AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389
 DB 1090 GCCACGCGGAGTCAACCTGCTCTCATCAGCGCGTGCACCTACCGCCACACCTGCCCGAG 1149
 QY 390 HisThrAlaTrpValAsnTyrValArgSerHisAspAlleGlyTrpThrPheAlaAsp 409
 DB 1150 CATACCGCTGGGTCACTACGTCCGACGACGACGACATCGCTGGACGTTTGGCGAT 1209
 QY 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 429
 DB 1210 GAAGACCGGCATATCTGGGCATAAAGCGGTACGACCCGCGCAATTCCTCAACCGCTTC 1269
 QY 430 PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer 449
 DB 1270 TTCTCAACCTTTTCGACGGCAGCTTCGCTCGTGGCTACGCTTCCATACACCCCAAGC 1329
 QY 450 ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp 469
 DB 1330 ACAGCGACTCGCGTGTCAAGTGTACAGCGCGGCATTTGGTGGCTTGGCGCAAGACGAT 1389
 QY 470 ProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu 489
 DB 1390 CCCCACCGCGTGTACCGCATCAAACTCTTGTACAGCATTCCTTTAGTACCGCGGTCTG 1449
 QY 490 ProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspAspTrpSerGlnAsp 509
 DB 1450 CGCTGATTTACCTAGCGGACGAGTGGGTACGCTCAATCACGACGACTGGTGCAGAGAC 1509

QY 510 SerAsnLysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeuTyr 529
 DB 1510 AGCAATAAGACGACGACGAGCCGTTGGCGGACCGCTCGCGCTACCAAGACCGCTGTAC 1569
 QY 530 AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMet 549
 DB 1570 GCGCAACCAACAGATCCGTCACCGACGCGCGGCAATCTATCAGGGCTTGGCCCATATG 1629
 QY 550 IleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThr 569
 DB 1630 ATTGCGCTCCGCCAAGCAATCCGCTTCGACCGCGGACGCTGCTTACATTCACACACC 1689
 QY 570 AsnAsnLysHisIleIleGlyTyrIleArgAsnAsnAlaLeuAlaPheGlyAsnPhe 589
 DB 1690 AACCAACAGACATCATCCGCTACATCCGCAACAAATGGCGCTTTGGCATTCGTAACCTTC 1749
 QY 590 SerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis 609
 DB 1750 AGCGAATATCCGCAACCGTTACCGCGCATACCTCTCAAGCCATCCCTTCAAGGCGCAC 1809
 QY 610 AspLeuIleGlyGlyIleValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
 DB 1810 GACCTCATCGGTGGCAAACTGTACGCTGAATCAGGATTTGACGCTTACGCCCTATCAG 1869
 QY 630 ValMetTrpLeuGluIleAla 636
 DB 1870 GTCATGTGCTCGAAATCGCC 1890
 RESULT 5
 LOCUS AX107016 6851 bp DNA linear PAT 30-APR-2001
 DEFINITION Sequence 7 from Patent WO0125449.
 ACCESSION AX107016
 VERSION AX107016.1 GI:13922550
 KEYWORDS synthetic construct
 SOURCE synthetic construct
 ORGANISM artificial sequences.
 REFERENCE 1
 AUTHORS Bengs, H., Polakowski, T., Held, A. and Gallert, K.C.
 TITLE Novel immobilizable amyloamylase, the use thereof and method for
 JOURNAL producing poly(1,4-alpha-glucan)
 Patent: WO 0125449-A 7 12-APR-2001;
 Axiva GmbH (DE)
 FEATURES
 Location/Qualifiers
 1..6851
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="Beschreibung der kunstlichen
 Sequenz: Expressionsvektor pGEX-4T-1-Amsus"
 BASE COUNT 1668 a 1791 c 1759 g 1633 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5,04e-282 Length: 6851
 Score: 3360.00 Matches: 626
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.84% Mismatches: 0
 Query Match: 98.59% Indels: 0
 DB: 6 Gaps: 0
 US-09-843-007A-2 (1-636) x AX107016 (1-6851)
 QY 10 LeuIleLeuGlnTyrIleuLysThrArgIleLeuAspIleTyrThrProGluGlnArgAla 29
 DB 945 ATGATTTTACGTACCTCAAAACACGCATCTTTGGACATCTACACGCCCGCCGACGCGCC 1004
 QY 30 GlyTleGluLysSerGluAspTrpArgGlnPheSerArgArgMetAspThrHisPhePro 49
 DB 1005 GGCATCGAAATATCCGAAGACTGGCGGAGTGTTCGCGCCGCGATGATGATGATTTCCCC 1064
 QY 50 LysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsnGluAlaLeuProMetLeu 69

1065 AAACTGATGAACGAACCTGACAGCGGTGACGCAACAAAGACCCCTGCTGCTATGCTG 1124
 70 GluMetLeuAlaGlnAlaTirpGlnSerTyrSerGlnArgAsnSerSerLeuIlyAsp 89
 1125 GAAATGCTGCTGGCGAGCATGGCAAGCATTTCCCAAGCAACTCATCTCTTAAAGAT 1184
 90 IleAspIleAlaArgGluAsnProAspTirpIleLeuSerAsnIlyGlnValGly 109
 1185 ATCGATATGCGCGCGGAAACAAACCCGATTCGATTTTGTCCAAACAAACAAATCGCGCGC 1244
 110 ValCysTyrValAspLeuPheAlaGlyAspLeuIlyGlyLeuIlyAspIlyProTyr 129
 1245 GTGGCTACCTGATTTGTTTCCGCGCATTTGAAGGCTTGAAGATTAATTCCTTAT 1304
 130 PheGlnGluLeuGlyLeuThrTyrIleuHisLeuMetProLeuPheIlyCysProGluGly 149
 1305 TTTCAGAGCTTGGTTTGAATTCATTCGACACCTGATGCGCTGTTTAAATGCTCGAAGC 1364
 150 LysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThr 169
 1365 AAAGCGAGCGCGGTATTCGCTCAGCAGCTACCGCATTCATCCGCGCATGGGCA 1424
 170 IleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIleSerAlaVal 189
 1425 ACAGCGACTTGGCGCAAGTATTCGCTGCTGCAAGACCGCGCATTTCCGCGCTGTC 1484
 190 AspPheIlePheAsnHisThrSerAsnGluHisGluTirpAlaGlnArgCysAlaGly 209
 1485 GATTTTATCTTCACCCACACCTCCACGAAACGAATGGCGCAACGCTGCGCGCGCGC 1544
 210 AspProLeuPheAspAsnPheTyrIlePheProAspArgArgMetProAspGlnTyr 229
 1545 GACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCCAATAC 1604
 230 AspArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGlu 249
 1605 GACCGCACCTCGCGAAATCTTCCCGACGACGACCGCGCGCTTCTCGCAACTGGAA 1664
 250 AspGlyArgTirpValTirpThrPheAsnSerPheGlnTirpAspLeuAsnTyrSerAsn 269
 1665 GACGAGCTGGGTGGGAGCAGCTTCATTTCTTCATTCAGTGGACTTGAATACAGCAAC 1724
 270 ProTirpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
 1725 CCGTGGGTATTCGCGCAATGGCGCGAAATGCTGCTTCTCCCAACTTGGCGCTGAC 1784
 290 IleLeuArgMetAspAlaValAlaPheIleTirpIlyGlnMetGlyThrSerCysGluAsn 309
 1785 ATCTCGGTATGGATCGGTGCTTTATTTGGAACAAATGGGCAAGCTGCGAAGAC 1844
 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro 329
 1845 CTGCGCAGCGCGCACCCCTCATCCGCGCTTCAATGCGTTATGCGTATTGCGCGGCC 1904
 330 AlaValPhePheIlySerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349
 1905 GCGGTGCTCTCAAAATCCGAAGCATTCGTCACCCCGACCAAGTCTCCAAATACATCGGG 1964
 350 GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTirpAsnThrLeu 369
 1965 CAGGAGCAATGCCAATTCGTTNACACCCCTCGCAATGGCATTTGTGGACACCCCTT 2024
 370 AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389
 2025 GCCACGCGGGAATCAACCTGCTCCATCAGCGCTGACCTACCGCCACCAACCTGCCGAG 2084
 390 HisThrAlaTirpValAsnTyrValArgSerHisAspAllyGlyTirpThrPheAlaAsp 409
 2085 CATACCGCTGGGTCAACTACGTCCGACGACGACATCGCTGGACGTTTGGCAT 2144
 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 429

2145 GAAGACGGCGCATATCTGGGCATTAAGCGGTACAGCACCCCAATTCCTCAACGCTTC 2204
 430 PheValAsnArgPheAspGlySerPheAlaArgGlyValPropheGlnTyrAsnProSer 449
 2205 TTGTCACACCGTTTCGACGGCAGCTTCGCTCGTGGCGTACCGTTCCAATCAACCCAGC 2264
 450 ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp 469
 2265 ACAGCGACCTCGGTGTCAGTGGTACAGCGCGGCATTTGGCTGGCCAGACCAT 2324
 470 ProHisAlaValAspArgIleLeuLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu 489
 2325 CCCACGCGGTGACCGCATCAAACTCTTGTACAGCATGCTTTGAGTACCGCGGCTG 2384
 490 ProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspAspTirpSerGlnAsp 509
 2385 CCGTGTATTTACCTAGCGAGCAAGTGGGTACGCTCAATGACGACGACTGGTTCGCAAGAC 2444
 510 SerAsnIlySerAspAspSerArgTirpAlaHisArgProArgTyrAsnGluAlaLeuTyr 529
 2445 AGCATAAGACGAGCAGCAGCGGTGGCGCACCGTCCGCGCTACACGAAAGCCCTGTAC 2504
 530 AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyIleuArgHisMet 549
 2505 CGCAACGCAACGATTCGCTCGACCGCGCGCGGCAATCTATCAGGCTTTCGCGCATATG 2564
 550 IleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThr 569
 2565 ATTCCGCTCCGCCAAGCAATCCGCGCTTCGACGGCGGCGCATTCGTTACATTCACAC 2624
 570 AsnAsnIlyHisIleIleGlyTyrIleArgAsnAsnAlaLeuLeuAlaPheGlyAsnPhe 589
 2625 AACCAACGACCATCATCGCTACATCCGCAACAAATGCGCTTTTGGCATTCGGTAATTC 2684
 590 SerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheIlyAlaHis 609
 2685 AGCGAATATCCGCAACCGTTACCGCGCATCCCTTCGACGCGCATCCCTTCAGCGCGCAC 2744
 610 AspLeuIleGlyGlyTyrThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
 2745 GACCTCATCGTGGCAAACTGTCAGCTGATCAGGATTTGCGCTTCAGCGCTTATCAG 2804
 630 ValMetTirpLeuGluIleAla 636
 2805 GTCATGTGGCTCGAAATCGCC 2825

RESULT 6
 AY099334 1967 bp DNA linear BCT 13-JAN-2003
 Locus Neisseria meningitidis strain 93246 amylosucrase Ams (ams) gene,
 DEFINITION complete cds.
 ACCESSION AY099334
 VERSION AY099334.1 GI:27728141
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis
 Neisseria meningitidis
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 1 (bases 1 to 1967)
 Zhu, P., Teang, R.-S.W. and Tsai, C.M.
 Nonencapsulated Neisseria meningitidis Strain Produces Amylopectin
 from Sucrose: Altering the Concept for Differentiation between N.
 meningitidis and N. polysacchara
 J. Clin. Microbiol. 41 (1), 273-278 (2003)
 12517860
 2 (bases 1 to 1967)
 Zhu, P.
 Direct Submission
 Submitted (23-APR-2002) LBF/DBPAP, CBER, 8800 Rockville Pike,
 Bethesda, MD 20892, USA
 Location/Qualifiers
 1..1967
 /organism="Neisseria meningitidis"

/mol_type="genomic DNA"		101 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis		200
/strain="93246"		561 CACGAAGCGCGCATTTCCGCGCTGCGATTATTTATCTTCAACACACATCTCCAAACGACAC		620
/db_xref="taxon:487"		201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe		220
21..1931		621 GAATGGCGCAACGCTGCGCGCCGCGACCGCTTTTCGACAAATTTCTACTATATTTTC		680
/gene="ams"		221 ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln		240
21..1931		681 CCGCAGCGCGGATGCCCGACCAATACGACCGCACCTTCCGCGAAATCTTCCCGCAGCAG		740
/gene="ams"		241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrThrPheAsnSer		260
/codon_start=1		741 CACCGCGCGCGCTTCTCCAACTGGAAGACGACGCTGGGTGTGGACGACGTTCAATTC		800
/product="amylomucase Ams"		261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheAspGalaMetAlaGlyGluMet		280
/protein_id="AAW51152.1"		801 TTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCCGCGCAATGGCGGCGAATG		860
/db_xref="GI:27728142"		281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp		300
/translation="MLTPQQVGLIIQLKTRLDIYTPQRAGIEKSEDMRQFSRM		861 ATGTTCTCTTGCCAACTTGGCGGTTCACATCTTGGCTATGGATCGCGTTTTCCTTTATTGG		920
DTHPKMLNEELSVYGNKALLPMLELLAQWQSQRSAKLNKIDIERNNPWIL		301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe		320
SNQVGGCYVDLFDGLDKDKHIFGLTYLHLMPLFKPGKSDGDAYVSSY		921 AAACAATGGGACCAAGCTGGAAACCTGCCGAGGCGCACGCCCTCATCGCGCGTTC		980
RVDNVALGIDLRVIAALHEAGISAVVDFFPNHTSNEHEWAQRCAAGDPLDFNFFY		321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis		340
IFPDRMPQYDRLREIIPDHPGGSQLEDRGWMTTFNSFOMDLNYSNFWVRAM		981 AATGCGGTATGCGTATCGCGCGCGCGCGCTGTTCTTCAAAATCCGAAGCCATCGTCCAC		1040
AGSMPLANEGLVILMDAVAFIWKMGTSCEMLPOAHALIRAFNAVMISIAAPAVFFK		341 ProAspGlnValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu		360
SEALVHPDQVQYVIGDEQCIQYNPLQMALIWNLTATREVNLLHQALTYRNLPSHTA		1041 CCGACCAAGTCGTCCAATATATCGGCGAGGACGAATGCCAAATCGCTACAAACCCCTG		1100
WNYVPSHDDIGWTFADEDAAYILGICYSYDERQFLARFFVNRDGSFARGVPEQYDNT		361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla		380
GCKVSGTSLAAGLAQNDPFAVSRINKLIYSIALSTGGGLPIYIGDEVGTFPDGQVQ		1101 CAATGGCATTTGTGTGAACACACCTTCCACCGCGCGAAGTCAACCTGCTCCATCAGGCG		1160
DSNKSDDSRWAHPRNEALYACRNDPSTAAQIYQGLRMIYAVRQSNPRFDDGRLVT		381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis		400
FNTNNKHIIYTRNALLAFGNFSEYPTVTATHTLOAMPFKADHLIGGKTVSLNODLT		1161 CTGACCTACCGCCACCACTCGCCGAGCATATCTGCTGGGTCAACTAGCTCCGCGAGCCAC		1220
LOFYQVWLEIA"		401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr		420
BASE COUNT 463 a 626 c 490 g 388 t		1221 GACGACATCGGCTGACCGCTTTCGCGATGAAGCGGSCATATCTGGGCAATAGCGGTAC		1280
ORIGIN		421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg		440
Alignment Scores:		1281 GACCAACCGCCAGTTCTTCAACCGCTTCTTCGTCAACCGTTCGACGCGGCGCTTCGCGCGC		1340
Pred. No.: 7,17e-281		441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla		460
Score: 3338.00		1341 GCGGTACCTTCCAAATACACCCCAACGCGGCGACTGCGCGGTACGCGGTACGTCGCC		1400
Percent Similarity: 98.90%		461 AlaLeuValGlyLeuAlaGlnAspProHisAlaValAspArgIleIleLysLeuLeuTyr		480
Best Local Similarity: 97.80%		1401 GCATTTGGCAGGCTTGCTCAAAATGACCCCAACCGCGCTCAGCGCGCATCAAACTTTGTAC		1460
Query Match: 97.95%		481 SerIleAlaLeuSerThrGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr		500
DB: 1		1461 AGCATCGCATTTGAGTACCGGCGGGCTGCCCTTGATTTTACCTTCGCGCAAGAGTGGGACG		1520
US-09-843-007A-2 (1-636) x AY099334 (1-1967)		501 LeuAsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis		520
Qy 1 MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu		1521 CCCAACGACGACGCGGTGGGTGCAACAGCAGCAACAAAGACGACGACGACGCGTGGCGGAC		1580
Db 21 ATGTTGACCCCAACGACGCAAGTCGGTGTGATTTTACAGTACCTCAAAACACGCACTTG		521 ArgProArgTrpAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly		540
Qy 21 AspIleTyrThrProGlnGlnArgAlaGlyIleGlyLysSerGluAspTrpArgGlnPhe		1581 GTCGCGCGCTATACGAAGACCCCTGTACGCGCAACGACGATCCGTCGCGCGACCGCGG		1640
Db 81 GACATCTACACGCGCGAAGCGCGCGCATCGAATAATCGAAGACTGGCGCGCAGTTT		541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp		560
Qy 41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly				
Db 141 TCGCGCGCGCATGATACCGCATTTCCCAAACTGATGACGAACTCGACGCGGTGTACGCGC				
Qy 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTrpGlnSerTyr				
Db 201 AACACGAAAGCCCTGTGCTGCTATGCTGGAATGCTGCGCGCGGCGATGGCAGAGCTAT				
Qy 81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrp				
Db 261 TCCCAACGACGCGCGCTCCCTGAAAACATGATGATCAACGCGAAGCAACCCCGATTGG				
Qy 101 IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu				
Db 321 ATTTTGTCAACAAACAAAGTTGGCGGTGTGTGTACGTTGATTGTTTTCGCGCGCATTTA				
Qy 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrIleuHisLeu				
Db 381 AAAGGCTGAAAGACAAATTCATTATTTTCAAGAGCTTGGTTTGACTTATCTGCACCTG				
Qy 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr				
Db 441 ATGCGCGCTGTTAAATGCGCTGAAGCAAAAGCGACGCGCGCTATCGCGTCAGCAGCTAC				
Qy 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu				
Db 501 CCGCATGTCAATCGCGCATGGGCACAAATAGGCGACTTGGCGAAGTCAATCTGCGCTG				

```

Db      1641 CAAATCTATCAGGCTTGGCCATATGATTCGCTCGCAAGCAATCGCGCTTCGAC 1700
Qy      561 GlyGlyValThrPheSerThrAsnAsnLysHisLeileGlyTyrLeuArgAsn 580
Db      1701 GCGGCGAGCTGGTACCTTCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1760
Qy      581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
Db      1761 AACGGCTTTTGGCATTCGGCAACTTCAGGAATATCCGAAACCGTTACCGCGATACC 1820
Qy      601 LeuGlnAlaMetProPheLysAlaHisAspLeuLeileGlyGlyLysThrValSerLeuAsn 620
Db      1821 CTGCAAGCCATGCTTTTCAAGCGCAGCACTCATCGCGGTAAACTGTACGCTGAAT 1880
Qy      621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluLeAla 636
Db      1881 CAGGATTTGACGCTTCAGCCCTATCAGGTCATGTCGCTCGAAATGCC 1928

RESULT 7
AY099335 1967 bp DNA linear BCT 13-JAN-2003
LOCUS Neisseria polysaccharea strain 85322 amylosucrase Ams (ams) gene,
DEFINITION complete cds.
ACCESSION AY099335
VERSION AY099335.1 GI:27728139
KEYWORDS Neisseria polysaccharea
SOURCE Neisseria polysaccharea
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
          Neisseriaceae; Neisseria.
REFERENCE 1 (bases 1 to 1967)
AUTHORS Zhu,P., Tsang,R.S.W. and Teai,C.M.
TITLE Nonencapsulated Neisseria meningitidis Strain Produces Amylopentin
        from Sucrose: Altering the Concept for Differentiation between N.
        meningitidis and N. polysaccharea
JOURNAL J. Clin. Microbiol. 41 (1), 273-278 (2003)
PUBMED 12517850
REFERENCE 2 (bases 1 to 1967)
AUTHORS Zhu,P.
TITLE Direct Submission
JOURNAL Submitted (23-Apr-2002) LBP/DBPAP, CSER, 8800 Rockville Pike,
        Bethesda, MD 20892, USA
FEATURES
     source
         1..1967
             /organism="Neisseria polysaccharea"
             /mol_type="genomic DNA"
             /strain="85322"
             /db_xref="taxon:489"
         21..1931
             /gene="ams"
         21..1931
             /gene="ams"
             /codon_start=1
             /transl_table=11
             /product="amylosucrase Ams"
             /protein_id="AA051153.1"
             /db_xref="GI:27728140"
             /translation="MLTPTQVGLILQHLKTRILLDIYTPBOAGIEKSEDRQPSRM
             DTHFPLKMLNLSVYTGNEALLPMLLQAQWQSIQSRASLKNKIDIERENPDWIL
             SNKQVGVYDLDPAGDLKLDKIPYQELGLTYLHMLPLFKCPGSKSDGGYAVSSY
             RDNVPAIGTIDDLREVI AALHEAGI SAVVDFIFNHTSNEHWAQRCAAGDLFDFNY
             I PDRMPDQYDRLREI PPDQHPGPGSLEDGRWWTTFNSFQWDLNYSNPWFRAM
             AGEMFLANLGVDIRMDAVAFWKMGTSCELEPOAHALIRAFNAVNRIRAPAVFPK
             SEAI VHPQVQVQIGQDBCOIGYNPLQWALLANTLAREVNLHQAULTVRLNPEHTS
             WNYVRSDDIGWTADADAAYLGISGVDRHQFLNRFVNRFDGSFARGVFPYNPNT
             GCRVSGSAALAGLAKQNDPHAVSRNKLITSTALSTGLPLIYLDGVTGENDQWAG
             DKGKSDSRHWRPRYNALYAQRNDPTAAGCIYQGLRHMIYAVRQSNPRFDGGLVT
             FNTNNKHGIYIRNALLAFNFSFPQVTAHTLQAMPFPKADHLIGKTVSLNQDLT
             LPQVQWMLEIA"
         459 a 631 c 489 g 388 t
BASE COUNT
ORIGIN

```

Alignment Scores:

```

Pred. No.: 1,95e-280 Length: 1967
Score: 3333.00 Matches: 620
Percent Similarity: 98.90% Conservative: 5
Best Local Similarity: 97.48% Mismatches: 7
Query Match: 97.80% Indels: 0
DB: 1 Gaps: 0

US-09-843-007A-2 (1-636) x AY099335 (1-1967)

Qy      1 MetLeuThrProThrGlnGlnValGlyLeuLeileGlyLeuGlnTyrLeuLysThrArgileLeu 20
Db      21 ATGTGACCCCCACGACAGCAAGTCGGTTGATTTTACAGCACCTCAAACCGCATCTTG 80
Qy      21 AspIleTyrThrProGluGlnArgAlaGlyLeileGluLysSerGluAspTyrArgGlnPhe 40
Db      81 GACATCTACAGCCCGAACACGCGCGCGCATCGAAAAATCCGAAGACTGSGCGGAGTTT 140
Qy      41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
Db      141 TCGGCGCGCATGATGATGATTTCCCAAAATGATGACGAACTCGACAGCGGTGTACGCG 200
Qy      61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTyrGlnSerTyr 80
Db      201 AACACAGAGCCCTCTGCTGCTATGCTGGAATGCTCTGGCGCAGCATGSGAGACTAT 260
Qy      81 SerGlnArgAsnSerSerSerLeuLysAspIleAlaArgGluAsnAsnProAspTyr 100
Db      261 TCCCAACGCGAGCGCTCCCTGAAAAACATTCACATCGAACGCGAAAAACAACCCGATTTG 320
Qy      101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
Db      321 ATTTTGTCAACAAACAAAGTCGCGCGGTGTGCTGACGTGATTTGTTGCGCGCGGATTTG 380
Qy      121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
Db      381 AAGGCTTGAAGACAAATCCCTTATTTCAAGAGCTTGGTTGACTTATCTGACACTG 440
Qy      141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr 160
Db      441 ATGCGCTGTTTAAATGCCCTGAAGGCAAAACGCGCGCTATGCGGTGACGAGCTAC 500
Qy      161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
Db      501 CGCATGTCAATCCGCACTGGGCAATAGACAGCTTTCGCGAAGTATTTGCCGCGCTG 560
Qy      181 HisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHis 200
Db      561 CACGAAGCGGTATTTTCGCGCTGCTGATTTCTATCTTCAACCAACCTCCCAACGAAAC 620
Qy      201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
Db      621 GAATGGGCGCAACGCTGCGCGCGCGGACCCCGCTTTTCGACAAATTTCTACTATATTTC 680
Qy      221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
Db      681 CCGACGCGCGGATGTCGCGCAATACGACCCACCTCGCGCAAAATCTTCCCGGACCAAG 740
Qy      241 HisProGlyGlyPheSerGlnLeuAspGlyArgTyrValTyrThrPheAsnSer 260
Db      741 CACCGCGCGGCTTCTCGCACTGGAGACGAGCGCTGGGTGTGACGACGCTTCAATTC 800
Qy      261 PheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyLeuMet 280
Db      801 TTCCAATGGGACTTGATTAACAGCAACCCGCGGTATTTCGCGCAATGCGCGGCAAAATG 860
Qy      281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr 300
Db      861 CTGTTCTCTGCAACTTGGCGGTTCGACATCTTGGCTATGGAATGCGGTGTCGCTTTATTTC 920
Qy      301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuLeuArgAlaPhe 320
Db      921 AARCAATGGGTACAGCTCGAAAACTTCGCGAGGCGCACCGCTCATCCGCGGCTTC 980

```


QY	321	AspAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340
Db	981	AATGCGGTATGCGGTATGCGCGCGCGCGCGGTCTTCAATCCGAAGCCATCGTCCAC	1040
QY	341	ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360
Db	1041	CCGACCAAGTCTCCAAATATATCGGCGAGGAGCAATGCCAATCGCTACACCCCTTG	1100
QY	361	GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla	380
Db	1101	CAATGGCATTTGTGTGAACACCTTGCACCGCGGAACTCACTGCTCCATCAGGCG	1160
QY	381	LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis	400
Db	1161	CTGACTACCGCAACACCTGCGCGAGCATCTCTCGGTCACTACGCTCGCAGCCAC	1220
QY	401	AspAspIleGlyTyrThrPheAlaAspGluAspAlaTyrLeuGlyIleSerGlyTyr	420
Db	1221	GACCATCGGCTGGACGTTTCGCGATGAATGCGGATATCTGGGCATAGCGGCTAC	1280
QY	421	AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg	440
Db	1281	GATCACCGCAATCTCTCAACCGCTTCTTGTCAACCGTTTCAGCGGAGCTTGGCGCG	1340
QY	441	GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla	460
Db	1341	GGCGTACCTTCCAAATACACCCCAACACCGGGGACTGCGCGGTCAAGCGTAGCTCCG	1400
QY	461	AlaLeuValGlyLeuAlaGlnAspProHisAlaValAspArgIleLeuLeuTyr	480
Db	1401	GCATTCGCGGCTTGGCTCAAAATACCCCGCGCGCTCAGCGCATCAAACTTTGTAC	1460
QY	481	SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500
Db	1461	AGCATCGCATTGAGTACCGCGGCGCTGCGCTGATTTACTCTCGGACGAGCTGGCAG	1520
QY	501	LeuAsnAspAspTyrPheSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis	520
Db	1521	CCCAACGACGCGCGTGGCGGCAAGACGCAACAAAGGACGACAGCGGTGGGCGCAC	1580
QY	521	ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540
Db	1581	CGTCCGCGCTACACAGCGCTGTACGCGCAACGACGATCGTCCGACCGCGCGG	1640
QY	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
Db	1641	CAATCTATCAGGCTTGGCGCATATGATTTGCGTCCGCGCAAGCAATCCGCGCTTGCAC	1700
QY	561	GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleGlyTyrIleArgAsn	580
Db	1701	GGCGGCGGCTGTACTTTCAACACCAACCAACCAACATCATCGCTACATCCGCAAC	1760
QY	581	AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr	600
Db	1761	AACGCGCTTTTGGCATTCGCAACTTCAGCGAATATCCGCAACCGTTACCGCGCATACC	1820
QY	601	LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyValThrValSerLeuAsn	620
Db	1821	CTGCAAGCCATGCTTTTCAAGCGCGACGACCTCATCGCGCGCAAACTGTACGTTGAAT	1880
QY	621	GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla	636
Db	1881	CAGGATTTACGCTTCAGCCCTATCAGGTATGCTGCTGCGCTCGAATCGCC	1928
RESULT 8			
A47369			
LOCUS			
DEFINITION			
Sequence 1 from Patent WO9531553.			
ACCESSION			
A47369.1			
VERSION			
A47369.1			
KEYWORDS			
Neisseria polysaccharea			
SOURCE			
Neisseria polysaccharea			
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;			
Neisseriaceae; Neisseria.			
REFERENCE			
1 (bases 1 to 2883)			
AUTHORS			
Kosmann, J., Buettcher, V. and Welsh, T.			
TITLE			
DNA SEQUENCES CODING FOR ENZYMES CAPABLE OF FACILITATING THE			
SYNTHESIS OF LINEAR alpha -1,4 GLUCANS IN PLANTS, FUNGI AND			
MICROORGANISMS			
Patent: WO 9531553-A 1 23-NOV-1995;			
INST GENBIOLOGISCHE FORSCHUNG (DE)			
Other publication DE 447388 960627			
Other publication AU 2614195 951205.			
FEATURES			
Location/Qualifiers			
1..2883			
/organism="Neisseria polysaccharea"			
/mol_type="genomic DNA"			
/db_xref="taxon:489"			
/clone="PNE2"			
/clone_lib="GENOMIC LIBRARY IN PBLUESCRIPTII SK"			
939..2783			
/note="unnamed protein product"			
/codon_start=1			
/transl_table=11			
/protein_id="CAA03058.1"			
/db_xref="GI:2301370"			
BASE COUNT			
661 a 852 c 758 g 612 t			
ORIGIN			
Alignment Scores:			
Pred. No.: 1,58e-263 Length: 2883			
Score: 3141.50 Matches: 621			
Percent Similarity: 98.11% Conservative: 3			
Best Local Similarity: 97.64% Mismatches: 11			
Query Match: Indels: 9			
DB: Gaps: 1			
US-09-843-007A-2 (1-636) x A47369 (1-2883)			
QY	1	MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu	20
Db	939	ATGTTGACCCCGACGAGCAAGTCGGTTTGATTTTACAGTACCTCAAAACACGATCTTG	998
QY	21	AspIleTyrThrProGluGlnArgAlaGlyIleGlyLysSerGluAspTyrArgGlnPhe	40
Db	999	GACATCTACACGCCGACAGCGCGCGCATCGAAAAATCCGAAGACTGCGCGCAGTTT	1058
QY	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
Db	1059	TCGCGCGCGCATGGATACCGCATTTCCCAAACTGTATGAACGAACTCGACAGCGTGACG	1118
QY	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTyrGlnSerTyr	80
Db	1119	AACAAACGAGCCCTGCTGCTATGCTGAAATGCTGCTGCGCGCAGGCATGCAAGCTAT	1178
QY	81	SerGlnArgAsnSerSerLeuLysAspIleAlaArgGluAsnAsnProAspTyr	100
Db	1179	TCCCAACGCAACTCATCTTAAAGATATCGATATCGCGCGCAAAACCAACCCGATG	1238
QY	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
Db	1239	ATTGTTCCCAACAAACGATCGCGCGGTGTGCTACGTTGATTTGTTTGGCGCGATTG	1298

QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
 DB 1299 AAGGGCTTGAAAGATAAAATTCCTTATTTCAAGAGCTTGGTTGACTTATCTGCACCTG 1358
 QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
 DB 1359 AGCCGCTGTTTAAATGCCCTGAAGCAAAAGCGGCGGTATGCGGTACAGCAGTAC 1418
 QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
 DB 1419 CGCGATGTCAATCCGGCACTGGGCACAATAGCGGACTTGGCGGAAGTCAATTGCTGGCTG 1478
 QY 181 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200
 DB 1479 CACGAATGCG--AATTGCGCGCTGTGCTGATTTATCTTCAACCAACACCTCCCAACGAACAC 1536
 QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAspPheTyrTyrIlePhe 220
 DB 1537 GAATGGCG-CAACGCTGC---GCCGCGCGCCGCTTCTTGACAATTTCTACTATATTTTC 1592
 QY 221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
 DB 1593 CCGGACCGCGGATGCCGACCAATACGACCCGCTGCGGAAATCTTCCCGGACGAC 1652
 QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTrpThrPheAsnSer 260
 DB 1653 CACCCGCGCGCTTCTCGCACTGGAAGACGACCTGGGTGTGGACGACCTTCAATTC 1712
 QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280
 DB 1713 TTCCAATGGGACTTGAAATTACAGCAACCCGCTGGGTATTTCG- GCAATGGCGGGGAAATG 1771
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300
 DB 1772 CTGTTCCTTGCCAACTGGCGGTGACATCTCGGTATGGATGGGTTGCGCTTTATTTGG 1831
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
 DB 1832 AAACAAATGGGACAGCTGGAAACCTGC--GCAGCGCAGCGCTCATCGCGGCTTC 1889
 QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
 DB 1890 AATGCGGTATGCGTATTGCGCGCGCGCGCTGTTCTTCAAAATCCGAAGCCATCGTCCAC 1949
 QY 341 ProAspGlnValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
 DB 1950 CCGGACCAAGTCGTCCTAATACATCGGCGAGGACGAATGCCAAATCGGTACACCCCTG 2009
 QY 361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380
 DB 2010 CAAATGGCATTTGTTGGAAACACCTTGCCACGCGCGAAGTCAACCTGCTCCTCAACAGCG 2069
 QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400
 DB 2070 CTGACTTACCGCCCAACCTGCCCGAGCATACCGCTCGGTCACTAGTCCGCGACGCCAC 2129
 QY 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr 420
 DB 2130 GACGACATCGGCTGAGCGTTTGGCGATGACAGCGCGCATATCTGGGCAATAGCGGCTAC 2189
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
 DB 2190 GACCACCGCCCAATTCCTCAACCGCTTCTTCGTCAACCGTTTCGACGCGGACGCTCGCTCGT 2249
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
 DB 2250 GCGGTACCGTTCCTCAATCAACCCAGCACAGCGGACTGCCGTGTCTGAGTGTACAGCGCG 2309
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480
 DB 2310 GCATTTGGTTCGCGTGGCGCAAGCAGCATCCCGCGCGTGTACCGCATCAAACTCTGTGAC 2369
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500

DB 2370 AGCATTTGCTTTAGTACCGCGGCTCTCCGCTGATTTTACCTAGCGACCAAGTGGGTACG 2429
 QY 501 LeuAsnAspAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis 520
 DB 2430 CTCATGACGACGACTGCTGTCGAAGC-AGCAATAGAGCGGACGACGCGCTTGGGC-CAC 2487
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
 DB 2488 CGTCGCGCTACAAACGAAGCCCTGTACGCGCAACGGAAGCATCTCGACCGACGCGGC 2547
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 DB 2548 AA-ATCTATCATGGGCTTTCGCCCATATGATTCGCTCGCCAAAGCAATTCGCGCTTCAC 2606
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
 DB 2607 GCGCGCAGGCTGGTTACATTCACACCAACACACACATCATCGGTACAT-CGCAAC 2665
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
 DB 2666 AATCGCTTTTGGCATTCGTTAATTCAGCGCAATATCGCAAAACGTTTACCGCATACC 2725
 QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
 DB 2726 CTGCAAGCCATGCGCTTCAGGCGCACGACTCATCGTGGCAAAACTGTTCAGGCTGAT 2785
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
 DB 2786 CAGGATTTGAGCTTCAGCCCTATCAGTCTATGCTGCTCGAATTCGCC 2833
 RESULT 9
 A47785 LOCUS A47785 2883 bp DNA linear PAT 07-MAR-1997
 DEFINITION Sequence 1 from Patent WO9533838.
 ACCESSION A47785
 VERSION A47785.1 GI:2301671
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria polysaccharea
 Neisseria polysaccharea
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 Kossman,J., Buetcher,V. and Welsh,T.
 MICROORGANISMS PERMITTING THE INTRACELLULAR POLYHYDROXY ALKANATE
 SYNTHESIS WITH SIMULTANEOUS EXTRACELLULAR POLYSACCHARIDE SYNTHESIS
 AND PROCESSES FOR PRODUCING THE SAME
 Patent: WO 9533838-A 1 14-DEC-1995;
 INST GENBIOLOGISCHE FORSCHUNG (DF)
 Other publication AU 2787895 960104
 Other publication DE 4420223 950504.
 FEATURES
 source
 1. .2883
 /organism="Neisseria polysaccharea"
 /mol_type="genomic DNA"
 /db_xref="taxon:489"
 /clone="PMB2"
 /clone_lib="GENOMIC LIBRARY IN PLBUESCRIPTII SK"
 939, 2783
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA03083.1"
 /db_xref="GI:2301672"
 CDS
 translations="MLTPTQVGLILQYLKTRILDTYTPQRAGIEKSDWROFSRSM
 DTHFKLMVELDSVYGNNEALPMLEMLAQWQSYORSNLSKLDIDIAENPNPWTL
 SNKQVGVGYDLFAGDLGKDKLKYFOELGTYLEMLPKCPKSGPGGYAVSTY
 RDNYPALGTDLREVLALHSHFRFRFYLPQLQRTMAORCAGDPLDFNYLIF
 EDPMSPQDRTLRIFPDQFGGFSQLEDGRVWVTFNFSQMDLNTSNPWFAQWRA
 KCCSLTVALTSCVWRLPLGNKGQAARTCAHALIRAFNAWMIAPAFVFSSEA
 IVPDQVQVIGDEQIGYNPLQWALNTLALNTLHLYTHLHPHTAWN
 YVSHDDIGWTFADDEAAVLGSCYDHRQFLNFFVNRPDGTGARGVPQYNSTGDC
 RVSGTAALVGLAQDDPHAVDRIKLLYSIALSTALSTGLPLVLVSGDVTLNDDDWCAAI

RATTAVGPSALQORSVPRATERSVRSROIYQGLRHMTAVRQSNRPFQGGRLVTFNTN
NKEIIGYIATMPFWSVTSANIRKXULPRPCPSRRITSSVAXLSA"
BASE COUNT 661 a 852 c 758 g 612 t
ORIGIN

Alignment Scores:

Pred. No.: 1.58e-263 Length: 2883
Score: 3441.50 Matches: 621
Percent Similarity: 98.11% Conservative: 3
Best Local Similarity: 97.64% Mismatches: 11
Query Match: 92.18% Indels: 9
DB: 6 Gaps: 1

US-09-843-007A-2 (1-636) x A47785 (1-2883)

QY 1 MetLeuThrProThrGlnGlnValGlyLeuLeuLeuGlnInTyLeuLysThrArgIleLeu 20
DB 939 ATGTTGACCCCAAGTGGTGTGATTTTACAGTACTCAAAACAGCATCTTG 998
QY 21 AspIleTyrThrProGlnGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe 40
DB 999 GACATCTACAGCCCGGAAAGCGCGCGCATCGAAAAATCCGAAGACTGGCGCGCAGTTT 1058
QY 41 SerArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
DB 1059 TCGCGCGCATGGATACGATTTTCCCAAACTGATGAAAGAACTGGACAGCGTGTAGCGC 1118
QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr 80
DB 1119 AACACGAAGCCCTGCTGCTTATGCTGGAAATGCTGCTCGCGCGCATGGCAAGCTAT 1178
QY 81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnProAspTrp 100
DB 1179 TCCCAAGCACTCATCTCTTAAAGATATCGATATCGCGCGGAAACCAACCCGATTTG 1238
QY 101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
DB 1239 ATTTTGTCAACAAACAACTGCGCGCGTGTGCTAGTTGATTTGTTGCGCGCGATTTG 1298
QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
DB 1299 AAGGCTTTGAAGATTAATTCCTTATTTTCAGAGCTTGGTTGATTTATCTGCAOCTG 1358
QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
DB 1359 ATGCGCTGTTTAAATGCGCTGAAGCAAAAGCGACGCGGTATGCGGTGACGACGTAC 1418
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
DB 1419 CGCGATGTCATCCGCACTGGSCACATAGGCACTTCGCGCACTCATTTGCTGCGCTG 1478
QY 181 HisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSerAsnGluHis 200
DB 1479 CACGAATCGC--ATTTTCGCGCGTGTGATTTTATCTCAACACACACCTCCCAACGAACAC 1536
QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
DB 1537 GAATGCGC-CAGCGCTGC--GCGCGCGACCCGCTTTTCGCAATTTCTACTATATTTTC 1592
QY 221 ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
DB 1593 CCGACCGCGGATGTCGCCCAATACGACCGCACCTTCGCGGAAATCTTCCCGCACAG 1652
QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTrpThrThrPheAsnSer 260
DB 1653 CACCGCGCGGCTTCTCGAACTGGAGACGGACGCTGGGTGGGACGACCTTCAATATTC 1712
QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet 280
DB 1713 TTCCATGGGACTTGATTTACAGCACACCGTGGGTATTCGC-GCAATGCGCGGCGAATG 1771
QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValaPheIleTrp 300

DB 1772 CTGTTCTCTGCCAACTTGGCGTTGCATCTCTGCGTATGGATCGGTTGCCCTTTATTGG 1831
QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
DB 1832 AAACAAATGGGACACAGCTGGAAAACTGCG--CGAGCGCACGCCCTCATCGCGGCTTC 1889
QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
DB 1890 AATCGCGTATGCGTATTTGCGCGCCCGCGCTGTTCTTCAAATCGAAGCCATCGTCCAC 1949
QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
DB 1950 CCGACCAAGTCTGCTCAATACATCGGCGACGAGATGCCAAATCGGTATACACCCCTG 2009
QY 361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380
DB 2010 CAAATGGCATTTGTGTGAAACACCTTTGCCAGCGCGGAAGTCAACCTGCTCCATCAGCG 2069
QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400
DB 2070 CTGACCTACCCGCCACACCTGCTCCGAGCATACCGCTGGGTCAACTAGTCCGAGCCAC 2129
QY 401 AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420
DB 2130 CACGACATCGGCTGGAACCTTTGCGGATGAAAGCGCGCATATCTGGGCATTAAGCGGTAC 2189
QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
DB 2190 GACCACGCCAATTTCTCAACCGCTTCTGCTCAACCGTTTCGACGCGCGCTGCTCGT 2249
QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
DB 2250 GCGGTACCGTTTCCAATCAACCCCAAGCACAGCGGACTGCCGTGTCAGTGGTACAGCCGC 2309
QY 461 AlaIleValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480
DB 2310 GCATTTGGTGGCTTGGCGCAAGACATCCCGACCGCTTGACCGCATCAAATCTTTGTAC 2369
QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
DB 2370 AGCATTTGCTTTGAGTACCGCGGCTGCTGCGCTGATTTTACCTAGGCGACGAAGTGGTACG 2429
QY 501 LeuAsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis 520
DB 2430 CTCATGACGACGACTGGTGCCAAAGC-AGCAATAGACGACGACGCGTTGGGC-CAC 2487
QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspP-cSerThrAlaAlaGly 540
DB 2488 CGTCCGCGCTACACGAGCCCTGTACGCGCAACGGAAGATCCGTCGACCGCAGCCGCGC 2547
QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
DB 2548 AA-ATCTATACGGGCTTGGCCATATGATTCCCGTCCGCCAAAGCAATCCGCGCTTCGAC 2606
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
DB 2607 GCGCGCAGCTGGTTACATTCACACCAACCAACGACACATCATCGGTACAT-CGCAAC 2665
QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnInThrValThrAlaHisThr 600
DB 2666 AATGCGCTTTTGGCAATTCGTTAACTTCAGCGAATATCCGAAACCGTTACCGCGCATACC 2725
QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLeuThrValSerLeuAsn 620
DB 2726 CTGCAGCGCATGGCCCTTCAGCGCGCACACCTCATCGGTGGCAAACTGTACGCTGANT 2785
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
DB 2786 CAGGATTTGACGCTTCACGCCCTATCAGGTCTCATGTGGCTCGAAATCGCC 2833

RESULT 10

AE001946

LOCUS

10719 bp DNA linear

AE001946

BCT 22-NOV-1999

DEFINITION Deinococcus radiodurans R1 section 83 of 229 of the complete chromosome 1.
ACCESSION AE001946
VERSION AE001946.1
KEYWORDS GI:6458655
SOURCE Deinococcus radiodurans
ORGANISM Deinococcus radiodurans
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
Deinococcaceae; Deinococcus.
REFERENCE White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.
Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1
Science 286 (5444), 1571-1577 (1999)
JOURNAL 20036896
MEDLINE 10567266
PUBMED
REFERENCE White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischnann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct submission
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..10719
/organism="Deinococcus radiodurans"
/mol_type="Genomic DNA"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
72..1034
/gene="DR0932"
72..1034
/gene="DR0932"
/note="similar to SP:P19641 GB:X68873 PID:388220
PID:606125 GB:U00096 percent identity: 57.98; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polyprenyl synthase"
/protein_id="AAFI0509.1"
/db_xref="GI:6458656"
/translation="MPDAAPETRMREVLRSDFEITAGIEDIVTGGKRVPLVTLA
AALGAGDALATSHSWDELRLSYCVLLHSALLHDDLDSDKRGQSTAFRPG
NVVYMSGDVEMRLHLLSGMPSALTRAFQGAASVCEGVLFQVASYADYSFE
NYFQVHGKTAALLSLAQAPAMLLGADSDARDALSTGRYGHAFQMDLLDLLE
EAQIKGVGDREKATYFVLCLLGGPHDAEVALERRAGEDGVARMIALAKQR
TDDCTREIRREARALIAALDLPSPARTALAAERLENTH"
1109..13043
/gene="DR0933"
1109..13043
/gene="DR0933"
/note="similar to PID:2094767 percent identity: 60.15;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="alpha-amylase"
/protein_id="AAFI0510.1"
/db_xref="GI:6458657"
/translation="MLTDFDLARLAEFDDRDRAETFRRLRYGPELADNIRAVYGN
HADALIGELVEMLHAYHAPADLRDLERLLRDLDWQGMWGVYAYVDRFAGTLR
VGGERLGVLYLHLLPLRFDGNDGGYAVQDYSVRPDLGTFIDLSALARE
LRGRGSLDVLVNEAHEHVAATAGAAVYDYPHIFDRTPQDPANRGTLPFI
PDPAGNTFWNEAGGWTTFNRSQWDVWGNPAVFRFELDLITLAIENRGEVRLD
AIALFWKELGTDCCNQPEVRLHTRALRAATRVAPAVAKAAIVAPGDLITHLGSRD

HHGRVSDWAYHNSLWVQWSSLASRDTLLTAALAAFPKPTNTTWGVVVRCHDIGW
AIADDAARVGLSGFAHRLFLSDYSGFFPGSGFARGLVFQHHPTQGDRIISGTASLA
GLDLALETGDAERVNDALAILLHAYVLMGVGGVPLVYMGDELALNDDTDAFAVAHA
ADNRWVEPQMDWELVASAQADAATGQPVTPAGMPAGLHLLAVRRTPHLHASTES
PLESPDPQVLLIRRHHTGVLLQVYNFSEHHTFTPTWPLQEQLCAGVADHLLGESQFH
LGGDLALEPFRALWLVAGG"
3065..3838
/gene="DR0934"
3065..3838
/note="similar to GB:AE000657 percent identity: 72.64;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, MerR family"
/protein_id="AAFI0511.1"
/db_xref="GI:6458658"
/translation="MLVVCSTSLMGSDAKHPVYVISVAAILVDMKHPOTLRYERKGL
IRPSSGKTRLYSERDIEHLREIRRLTQELGVXAGVEVNRLLQHQDDLLQOEFEAE
IERIEDELRQARPAAPADPPDRPVVISIAAAILVDMKHPOTLRYERKQLIH
PGRSGKTRLYSERDIEHLREIRRLTQELGVXAGVEVNRLLQHQDDLLQOEFEAE
RLQDDLSRMTTWTLPAPSEAGPEEAAAADAGEDQN"
3816..4907
/gene="DR0935"
3816..4907
/note="similar to GB:D10483 SP:P05637 GB:X04711 PID:216474
PID:40919 percent identity: 55.48; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphatase, putative"
/protein_id="AAFI0512.1"
/db_xref="GI:6458659"
/translation="MPEKTRTEHQELSLTFLRGVRLREPRFIVSDALSRPSAAPPVL
CEVFLNALPALSPVPAPLWGVGVGADKRLALLAGLTHGHTWGAHL
VFLGVDYDRGPDGLVRLVQSLAEQALACGEVTALLGNHEVMLAALFREQDPRD
RFGFFYRNSGGLTDLERLAPGLDARLPIILAAQOGLLHADSPPVILGNDP
QEVNRYAEILLSPDPGNAPAFANFVRLKFAQDCTQARLLETYGGSELVGHGT
PTLLQSGQDLPFETVFLYADGLCLAVDSCHAYFPDAGFVARLGEAPGGAER
VLAELVLPFPFLDGLSLTATLRPSGEG"
4977..6017
/gene="DR0936"
4977..6017
/note="similar to GB:AL009126 percent identity: 61.90;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI0513.1"
/db_xref="GI:6458660"
/translation="MPILLDLHPDHPLEGRKOLLEWVVOGVTFDAMTNLPAGL
RAELSYQLNPREDIETVSHSGSVKYLITLNDGRQMEAYMPLDKTICVSMVG
KAPCAKATGKMGFGNLTGPEITGVQVLAVAGGEGISPREINLVFMGMGEALLNE
NSMAARLHPDGLMGSKRVTLSVIAKGRRLAEEDDLGKLSIALHAPDEET
QQIIPGANSIEBIMAAADYQAVTGRVTMEYTLERGVNDHLWQELIADRLGLV
SHVNLIPMNPDGSDPFVSSSEEQIQAIFYDALQDRGVDSVRSRSGKAGAACQGLAR
RPGATGEVGA"
6345..8135
/gene="DR0937"
6345..8135
/note="similar to PID:1652192 percent identity: 41.71;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="tetrapeptide repeat family protein"
/protein_id="AAFI0514.1"
/db_xref="GI:6458661"

```

NEFEALVRYNNPEPHFGILGLALYQLGDLRGATPFEGQFMQFAPQRYEGPYNLGVIA
TRERYSFADLRLYGEALTQAQSPAPAAQVLLRALATEQGRVXDYVALAATYADLRRA
DPENVEYFHPHQALYQAGVADALPVYAVLERKPSLQAALLADLYVAQGLPORA
VRELSAAARVLTGTRATLLRQSDLLAQNNLRGALNAASEARDRLPRAIIRE
AELLAQCGQPPAAVYRDALALPKSAPITQAAALYLEGNQYPERAAQAOALKLT
PDASTARALVQVAVYRGQVTOARTALNSSVAVEDADSLLLGLSYVALKDVAS
APVLSVESVLRPTPARQNLAAALATAYAEAEALLRGLVTVDEVKNAGSVLLGLS
QRAQRPDDARQFKAAGLSARAKDALX"
/gene="DR0938"
/length="9319"
/db_xref="GI:6458663"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI0516.1"
/db_xref="GI:6458663"
/transl_table="MTSARTPSQRPSPRRPAPGRWPELVAAATLVLAGVGLGS
LLGERAPAPVPCAGQPTTSIPVSGAVVADGSPOTAAQAQCTAGTKIPAPAAAP
AKTIPPVLPPEPRMPAPPTPRPEVESTTTAASPIQPPVTOAPATQTPPTPRAQA
PATQPPATPAPAPAPAPATPTEPATTEPAAPAGPTPAVAPSQRTPLASDRV
MLGTFGERALATAGVAGLGVTVYVADLGNQFVAGQFPFADEASGQQAADIRAY
ARAELYPFGQTLUNPAPAAAPAPTEPASPAPTPTPTAAAPAPAGPVTVQVGA
FNNVEGAQRFEQLRAQGFSPVNAPEYTKVTVLLGLPLTGLSDLTSTEGRLDAAGLDHF
RLR"
/length="10020"
/gene="DR0939"
/length="9316"
/db_xref="GI:6458662"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI0515.1"
/db_xref="GI:6458662"
/transl_table="MSRVADIPATVGRVVYTVIRVLEEAQNVLRASSGELARRAGV
TPFQKDLTYFGRGTRGIGYTVAVLERLRLALGLDQTNVTVVGMGRGLHAIANY
PGASDYQFQNVGFDVAPDVGVREVRGLTIQHSQFPFVASVAGTIPROVDMGLTVP
AEHAQAAQALVAGVGGILNFAPVLTQDLHLPLBFAFAPGRREVTVENVDFLAGWK
RLAFYMLGPQAGPAES"
/length="10017"
/gene="DR0940"
/length="10017"
/db_xref="GI:6458664"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI0517.1"
/db_xref="GI:6458664"
/transl_table="MTWKMPFPVLLGLLGGCTVGLSGRVNVGVVDGADAGSQV

```

Alignment Scores:

Pred. No.:	4,35e-104	Length:	10719
Score:	1319.50	Matches:	281
Percent Similarity:	58.62%	Conservative:	110
Best Local Similarity:	42.13%	Mismatches:	215
Query Match:	38.72%	Indels:	61
DB:	1	Gaps:	14

US-09-843-007A-2 (1-636) x AE001946 (1-10719)

QY	1	MetLeuThrProThrGlnValGlyLeuLeuGlnTyrLeuLysThrArgGlyLeu 20
Db	1109	ATGCTACGGCTGAC-----CTCGCCGCCCGCTG--- 1138
QY	21	AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTTPArgGlnPhe 40
Db	1139	-----CGCTCGCTTTCAGCATGACCGTGAAGCGCGAGACTTTC 1177
QY	41	SerArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60

Db	1178	COGCTCGCGTGGAGCGGTACGCGCCGGAACCTGGCGGCAACCTCGGCGGTCTACGGA 1237
QY	61	AsnAsn---GluAlaLeuLeu---ProMetLeuGluMetLeuLeuAlaGlnAlaTrrPglN 78
Db	1238	AACACGCGACGCGGTATAGTAGCGAGCTTCTAGAGTCAATGCTG-----CAC 1285
QY	79	SerTyrSerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluLeuAsnPro 98
Db	1286	GCTTACACCG 1345
QY	99	AspTrrIleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGly 118
Db	1346	GACTGGCTGCAAGGACCGAGATGGTGGCTAGTGGCCCTAGCTGACCGCTTGGCGGC 1405
QY	119	AspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeu 138
Db	1406	ACGCTCAGCGCGGTGGCGGCGGTGGAGTATCTGGAAGGGTGGCGCGTACCTACCTT 1465
QY	139	HisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSer 158
Db	1466	CACCTCTCGCGTCTCG 1525
QY	159	SerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAla 178
Db	1526	GATTACCGGAGGTGGCG 1585
QY	179	AlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsn 198
Db	1586	GAGCTCG 1645
QY	199	GluHisGluTrrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 218
Db	1646	GAACACGATGGCGGTGGCG 1705
QY	219	IlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluLeuPhePro 238
Db	1706	ATTTTCTCTGATCGACG 1765
QY	239	AspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTrrValTrrThr 257
Db	1766	GACTTCGCGCGGTGAATTTTCACTGGAACGGAGAGCGCGCGCGCGCGCGCGCGCGCG 1825
QY	258	PheAsnSerPheGlnTrrAspLeuAsnTyrSerAsnProTrrValPheArgAlaMeta 277
Db	1826	TTCAACCGCTCCAGTGGGACGTGAACCTGGGGAACCCCGCGCGTGTTCGCGGATATCTC 1885
QY	278	GlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAla 297
Db	1886	GACCTCATTTCTACCTTCGCAACCGGCGGCGTGGAGGTCTTCGGCTCAGCCATCGCT 1945
QY	298	PheIleTrrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIle 317
Db	1946	TTTTTGTGAACGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2005
QY	318	ArgAlaPheAsnAlaValMetArgIleAlaProAlaValPhePheLysSerGluAla 337
Db	2006	CACGCGCTCG 2065
QY	338	IleValHisProAspGlnValGlnTrrIleGlyGlnAspGlu----- 352
Db	2066	ATTGTGGCG 2125
QY	353	CysGlnIleGlyTrrAsnProLeuGlnMetAlaLeuLeuTrrAsnThrLeuAlaThrArg 372
Db	2126	TCCGACATGGCTTACCACAACTCGCTGATGGTGGAGCTCTGGAGAGCGCTCGCTCGGT 2185
QY	373	GluValAsnLeuLeuHisGlnAlaLeuThrTrrArgHisAsnLeuProGluHisThrAla 392
Db	2186	GACACGCGGTCTCTCACGCGCGCGCTGGCGCGCTTTCCTCCCAACGCGCGGACGACG 2245
QY	393	TrrValAsnTrrValArgSerHisAspIleGlyTrrThrPheAlaAspGluAspAla 412

Db 2246 TGGGGCGCTCTACGCGCGCTGTACGACGACATCGCTGGCGGATCGCGCGACGAGCGCG 2305

Qy 413 AlatyrluLeuglyleserGlyTyraSpHisarglnPheLeuAsnArgPhePheValAsn 432

Db 2306 GCGCGCTGGCGCTCTACG 2365

Qy 433 ArgPheAspGlySerPheAlaArgGlyValProPheGlnTyraAsnProSerThrGlyAsp 452

Db 2366 GAGTTTCCCGCGCTCTTTGGCGCGGAGTGGTGTTCAGCACCCACACCGCGCGCGAC 2425

Qy 453 CysArgValSerGlyThrAlaAlaLeuValGlyLeu----- 465

Db 2426 CGCGCGATTCG 2485

Qy 466 ---AlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyraSerIleAlaLeu 484

Db 2486 GACCGCGAGCGGTAACGACG 2545

Qy 485 SerThrGlyGlyLeuProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 504

Db 2546 GCCTTCGCGCGGCG 2605

Qy 505 AspTrpSerGlnAspSerAsnLysSerAspSerArgThrAlaHisArgProArgTyr 524

Db 2606 GACTTTCG 2665

Qy 525 AsnGluAlaLeuTyraAlaGln-----ArgAsnAspProSerThrAla 538

Db 2666 GACGCGGAGCGTGGTGGCTCTGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2725

Qy 539 AlaGlyGlnIleTyraGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg 558

Db 2726 GCGCGCGAAGTTCG 2785

Qy 559 PheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle----- 574

Db 2786 CTGCAAGCGC-----AGCACCGAGTCTGCTCCCTCCCGAGTCCCGAC 2827

Qy 575 -----IleGlyTyraIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsn 588

Db 2828 CCTTCGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2887

Qy 589 PheSerGlyTyraProGlnThrValThrAlaHisThrLeuGln---AlaMetProPheLys 607

Db 2888 TTCAGCGAGCACCATCATCTTCCCGAGCGTGGCGCTTGCAGAACAGCTCGCGCGCGCG 2947

Qy 608 AlaHisAspLeuIleGlyGlyLysThrValSerLeu---AsnGlnAspLeuThrLeuGln 626

Db 2948 GCCACGACCGCTGCTGCTGAGAGCGCGAGTTCACCTGGCGCGCGCGCGCGCGCGCG 3007

Qy 627 ProTyraGlnValMetTrpLeu 633

Db 3008 CGTACCGCGCGCTGTGGCTG 3028

RESULT 11

AE011997

LOCUS

DEFINITION

AE011997 10261 bp DNA linear BCT 29-MAY-2002

the complete genome.

AE011997 AE008923

VERSION

AE011997.1 GI:21109853

KEYWORDS

SOURCE

ORGANISM

Xanthomonas axonopodis pv. citri str. 306

Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

Xanthomonadaceae; Xanthomonas.

1 (bases 1 to 10261)

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,

Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R.H.A.,

Roski, A., Sana, J.A.D., Silva, C. de Souza, R.F., Spinola, L.A.F.,

Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezsa, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitajima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with

differing host specificities

Nature 417 (6887), 459-463 (2002)

2202145

22024217

2 (bases 1 to 10261)

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,

Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R.H.A.,

Roski, A., Sana, J.A.D., Silva, C. de Souza, R.F., Spinola, L.A.F.,

Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezsa, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitajima, J.P.

Direct Submission

Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de

Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,

Brazil

Location/Qualifiers

1. 10261

/organism="Xanthomonas axonopodis pv. citri str. 306"

/mol type="genomic DNA"

/strain="306"

/db xref="taxon:190486"

/notes="pathovar: citri"

155. 2629

/genes="fyua"

/notes="synonym: XAC3489"

155. 2629

/genes="fyua"

/notes="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/Genemark"

/codon start=1

/transl table=11

/product="TonB-dependent receptor"

/protein_id="AA038332.1"

/db xref="GI:21109853"

/translation="MRERLMSLTHTLWHLALACAVVSCICAPALAAQDRAAPATP

PAADSRVNDVSVFTGTSITATKLSXSVSVSTVGAEEVQSPAPRTAEIFRNIPGIR

SESSGEGNANIVRGLPVASGAKTLQDGLPVEFGDIAPGSDIFRSDFTILD

RIRAIKGSASTTINAPGGINFIKSTGTAGSGVSGRLDYDNTRIFNYGAPFA

EQVTFNIGGFRQDGVADGYTDDKGGQLKXNLTRLFHNGYVRYGKYLNDRAAGYL

PVPTSVGRGSPDLGCPGDPGNDTLYSNRTFDVLDENQVQPRETDLGDMHPIS

RTIGARAWFDLGNMLSEKFRVADNSGRFVSPPEVETDAASLAAISGGAGQIVOA

IDMDTWNSTVQSLGRNRLNVDVADGVSRSQNGLYAYGTFPFGDCCTVRSYRVD

VNAPYVLTDFSGKPSIDGSLRYDMGDARSINGTAIQAQLNDGVDGVIQVQEVQVAT

VTANSPRVYDNYLISYSLGGNYLITDILGAFARISGRANADRLFLFVGRDGSV

SSBEGVNVVQAEGAKWRDGLSLFATFSARTQENFEITSQRFYRQENYQVGL

EASVYRGFTLNGLTWDALISRDQITPENAGNVPRQADVVMQLTPSVYRGDNYQFG

VNLGCTDTAQANQLKMPGYQVNLFGDVRVTDALTVALNVNLFNTFGLTEAEEA

TIPIANGIIRARSAGRTTISLRYDP"

2664. 4598

/gene="XAC3490"

2664. 4598

gene

CDS

FEATURES

source

gene

CDS

gene

CDS

Db	3105	GTGGAAACCAAGCCTGGGCAGCAACGACGACTTGGTCGGCTTACACCGCCCTGCGGAA	3164
Qy	183	AleGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSerAsnGluHisGluTyr	202
Db	3165	GCGGCGATCAGCCTGTGGCTGACTTGGTGTCAATCACCGCGGACGATCAGCCTGG	3224
Qy	203	AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp	222
Db	3225	GCACAGCGCGCGCGCGCGGAGTACGACGTATCTCGACTACTACCACTACTTTGCGGAC	3284
Qy	223	ArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisPro	242
Db	3285	CGCACCTCGCCGACCGGTACGAAGCACCGCTGGCGGCGAGGTGTTCCGCATACAGCGCCC	3344
Qy	243	GlyGlyPheSerGlnLeuGluAsp--GlyArgTyrValTyrThrThrPheAsnSerPhe	261
Db	3345	GGCAACTTCACCTGGGTGGACGACACCGCGCAGTGGATGTGGACCAAGTCTTATCCGTAT	3404
Qy	262	GlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMetLeu	281
Db	3405	CAATGGGATTTGAATCTGGAGCAACCCGGCGGTGTTCCGGGCATCGGCCCTTGGCGATCGT	3464
Qy	282	PheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyrLys	301
Db	3465	CGGCTGGCCAACTCTGGCGTGGAAACATTCGCGCTGGATTCACCGGCTATCTGTGGAAA	3524
Qy	302	GlnMetGlyThrSerCysGluAsnLeuProGlnNlaHisAlaLeuIleArgAlaPheAsn	321
Db	3525	CGCATCGCACCGACGATTCGATGAATCAACCGGAGCGCACCCCTGCTGGTGGCGCTGCGC	3584
Qy	322	AlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisPro	341
Db	3585	GCGGTACCGATATCGTGGCGCGCGGTGTGTGATGAGGCGGAGCCATCGTGGCGATG	3644
Qy	342	AspGlnValValGlnTyrIleGlyGln-----AspGluCysGlnIleGly	356
Db	3645	ACGCAACTGCGCGCGCTACTTCGGCAGCGGGCTGTGATGAGGCGCCACGATGTCTATCGCC	3704
Qy	357	TyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeu	376
Db	3705	TACCACAGCAGCTGATGGCAGCGCGCTGTGTGGCACTGCGGTGTGCACCGCGGACATC	3764
Qy	377	LeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyr	396
Db	3765	CTGCACAACTGATCGGCGACAGCCCACTTGC CGCGCCACTGTGCATGCTGTGAGCTAT	3824
Qy	397	ValArgSerHisAspAlleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGly	416
Db	3825	GTGGGTGGCATGACGATATCGGTGGAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	3884
Qy	417	IleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGly	436
Db	3885	GCGCAACCGCGCTTTTCGCTCCGGGATGTGGCGGCTTCTATGTCACACGCGGTACCGGGC	3944
Qy	437	SerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGlyAsp-----CysArg	454
Db	3945	AGTTACGACGCGCGGAGATTTCCAG-----ACGACGCGGACGCGGTGCGATGGC	3995
Qy	455	ValSerGlyThrAlaAlaLeuValGlyLeu-----AlaGlnAspAspProHis--	471
Db	3996	ACCAACGCGCATGGTGTGTGCTGACTTCGCGGATCCAGCGCGGCGAGGCGGCGATGGC	4055
Qy	472	-----AlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThr	486
Db	4056	GCGCGGTGGCGGTGTGAGTGGACCGCGCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4115
Qy	487	GlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspTyr	506
Db	4116	CCGGCGGTTCGCTGATCTACATGGGCGCAATTTGGCGATGTGCATGATACCGCGCTAT	4175
Qy	507	SerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGlu	526

Db	4176	CGCCAGACCCGACCGCCAGCATGAAGCGCGTTGGCTGTCATCGCGCGCGATGGATGG	4231
Qy	527	AlaLeuTyxAlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrginglyLeu	546
Db	4236	CAGCTCCGCGCAGCGCTCAGATGCCACAGCCTGAGTGCACGGGTACCGACGCTTG	4295
Qy	547	ArgHisVetIlealavalArgGlnSerAsnProArgPheaspGlyGlyArg---	565
Db	4296	CGTGATGTGATCGCGCAACGTGCGCGTGTAGGCGCGTGGCAGCAGATCAAGCACTGC	4355
Qy	566	ThrPheAsnThrAsnAsnLysHisIleIleGlyTyxIleArgAsnAsnAlaLeuLeuAla	585
Db	4356	AGCATTCGCGCTGAATGATCTCGCGTGTCTGCTGACCGGTGCGACAGTTTCATCGCC	4415
Qy	586	PheGlyAsnPheSerGluTyxProGlnThrVal-----	598
Db	4416	TTGCATAACTTCACGACCGCGCTCGACGTGGAGCTTCGCGCTGCGTTGGCTGCAG	4571
Qy	599	HisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSer	618
Db	4476	TGGACGCTGCTTGCCATC-----GATGACGCGATTTGTTGGCGCGCGCGCAGG	4523
Qy	619	LeuAsnGlnAspLeuThrLeuGlnProTyxGlnValMetTrpLeuGlu	634
Db	4524	GCGCATGATCGATCGTGTCTTCGCTTCGCGCTTACGCGCTGCGTTGGCTGCAG	4571
RESULT 12			
LOCUS	AE012454	11605 bp	DNA linear BCT 23-MAY-2002
DEFINITION	Xanthomonas campestris pv. campestris str. ATCC 33913, section 362 of 460 of the complete genome.		
ACCESSION	AE012454	AE008922	
VERSION	AE012454.1	GI:21114602	
KEYWORDS	Xanthomonas campestris pv. campestris str. ATCC 33913		
SOURCE	Xanthomonas campestris pv. campestris str. ATCC 33913		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xanthomonas.		
REFERENCE	1 (bases 1 to 11605)		
AUTHORS	da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, I.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, P., Cardoso, J., Chamargo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.		
TITLE	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities		
JOURNAL	Nature 417 (6887), 459-463 (2002)		
MEDLINE	22022145		
PUBMED	12024217		
REFERENCE	2 (bases 1 to 11605)		
AUTHORS	da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, I.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, P., Cardoso, J., Chamargo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.		

Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinoia, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, P.F., Setubal, J.C. and Kitajima, J.P.

TITLE

JOURNAL
Direct Submission
Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900, Brazil

FEATURES

source
i. .11605

Location/Qualifiers
/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
/mol_type="genomic DNA"
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
164. .1357
/gene="oprO"
/note="synonym: XCC3353"
164. .1357
/gene="oprO"
/note="synonym: XCC3353"

CDS
/gene="oprO"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="porin O"
/protein_id="AA042623.1"
/db_xref="GI:21114603"
/translation="MINDTLRLRAITACAAICLAPAAHAADGEPVTAEGVRVWDF TVFNDERPTERTQFRLVWVAGFYGNFYAEAFAGLOVEPGSGILARDVY IAKTTSAGTLVQFQKPLDLDRTGNSYGPFLBERGYASTTLAPYKRAISWQTSRPD ATWSTAAYSLBISDNTAKGAGLGTLLTAPEMGDARLHLGVLAHRYSHPGSGGA SLLLRPPENDLANNRSLTLPAPADGRDITDEKSLVAVVWVGPNSQETFESEGLLD DGAQHAKVSLGWSWFTGESRGVDRKTRFSRIKTPNRPSPGAFELALYDNRGD QHLDGQPDINASTUSTLGGNWPYKPNLRVNLVINDSRNDRRTMNAVVDHTLAVTR FQYD"

gene
1582. .2910
/gene="citM"
/note="synonym: XCC3354"
1582. .2910
/gene="citM"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="Mg++ citrate complex transporter"
/protein_id="AA042624.1"
/db_xref="GI:21114604"
/translation="MTALGFGWITFMVIMSKRLSPVALITVPIVALLGGPGTG INEMLDGKKIAPTGMWMLFALYFGVMIDAGLPDPLVGRILRVKGDPLKIYMGTA ILAMLSLDGSGTMTITVSAMPLFYQRLGNALNTCTVTILASGVNMLTPWGQPTA RAATLHVPADVPFVPLPAMAMATAGIILALAWLGMERRRLGVRLPDPGNMLETSLPDDNALPRVDETDKPKLLVNLVLTALMAALVVGILPMPVLPMIGFALALMI NYPLAEORRLVNHAGVNLVSVSLIPAGITFGILSNTGMVEARSLFLAVIPDSWG PHLAVITLVSPPFTFSNADAPFYGVLPILSERAGHYGTPVEMARSLAQGPVHLL SPLEPSTVLVGLAKVDFAHQRFILKAWLVLSMLAGGLLGLFLPLAR"
3073. .3813
/gene="phbB"
/note="synonym: XCC3355"
3073. .3813
/gene="phbB"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="acetoacetyl-coA reductase"
/protein_id="AA042625.1"
/db_xref="GI:21114605"
/translation="MTLRLAYTSGVSGVGTALCKLARTGTTVAGCGPNSPKSAW LREQRELGFDFVASSGNADMDSTMAAFKAEVGEIDVLVNNAGSGRDLTFQMSR DDMNAIVASINLSTNTKQVVDGNTSRGWRIVNIGVSAHKQIGQINPATKAAV HGFSRLAQEASRGVTVNTITSPGVIASASISNFPDPVLDRLATVSPVVRIGKDAEVA

gene
1582. .2910
/gene="citM"
/note="synonym: XCC3354"
1582. .2910
/gene="citM"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="Mg++ citrate complex transporter"
/protein_id="AA042624.1"
/db_xref="GI:21114604"
/translation="MTALGFGWITFMVIMSKRLSPVALITVPIVALLGGPGTG INEMLDGKKIAPTGMWMLFALYFGVMIDAGLPDPLVGRILRVKGDPLKIYMGTA ILAMLSLDGSGTMTITVSAMPLFYQRLGNALNTCTVTILASGVNMLTPWGQPTA RAATLHVPADVPFVPLPAMAMATAGIILALAWLGMERRRLGVRLPDPGNMLETSLPDDNALPRVDETDKPKLLVNLVLTALMAALVVGILPMPVLPMIGFALALMI NYPLAEORRLVNHAGVNLVSVSLIPAGITFGILSNTGMVEARSLFLAVIPDSWG PHLAVITLVSPPFTFSNADAPFYGVLPILSERAGHYGTPVEMARSLAQGPVHLL SPLEPSTVLVGLAKVDFAHQRFILKAWLVLSMLAGGLLGLFLPLAR"
3073. .3813
/gene="phbB"
/note="synonym: XCC3355"
3073. .3813
/gene="phbB"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="acetoacetyl-coA reductase"
/protein_id="AA042625.1"
/db_xref="GI:21114605"
/translation="MTLRLAYTSGVSGVGTALCKLARTGTTVAGCGPNSPKSAW LREQRELGFDFVASSGNADMDSTMAAFKAEVGEIDVLVNNAGSGRDLTFQMSR DDMNAIVASINLSTNTKQVVDGNTSRGWRIVNIGVSAHKQIGQINPATKAAV HGFSRLAQEASRGVTVNTITSPGVIASASISNFPDPVLDRLATVSPVVRIGKDAEVA

gene
1582. .2910
/gene="citM"
/note="synonym: XCC3354"
1582. .2910
/gene="citM"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="Mg++ citrate complex transporter"
/protein_id="AA042624.1"
/db_xref="GI:21114604"
/translation="MTALGFGWITFMVIMSKRLSPVALITVPIVALLGGPGTG INEMLDGKKIAPTGMWMLFALYFGVMIDAGLPDPLVGRILRVKGDPLKIYMGTA ILAMLSLDGSGTMTITVSAMPLFYQRLGNALNTCTVTILASGVNMLTPWGQPTA RAATLHVPADVPFVPLPAMAMATAGIILALAWLGMERRRLGVRLPDPGNMLETSLPDDNALPRVDETDKPKLLVNLVLTALMAALVVGILPMPVLPMIGFALALMI NYPLAEORRLVNHAGVNLVSVSLIPAGITFGILSNTGMVEARSLFLAVIPDSWG PHLAVITLVSPPFTFSNADAPFYGVLPILSERAGHYGTPVEMARSLAQGPVHLL SPLEPSTVLVGLAKVDFAHQRFILKAWLVLSMLAGGLLGLFLPLAR"
3073. .3813
/gene="phbB"
/note="synonym: XCC3355"
3073. .3813
/gene="phbB"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="acetoacetyl-coA reductase"
/protein_id="AA042625.1"
/db_xref="GI:21114605"
/translation="MTLRLAYTSGVSGVGTALCKLARTGTTVAGCGPNSPKSAW LREQRELGFDFVASSGNADMDSTMAAFKAEVGEIDVLVNNAGSGRDLTFQMSR DDMNAIVASINLSTNTKQVVDGNTSRGWRIVNIGVSAHKQIGQINPATKAAV HGFSRLAQEASRGVTVNTITSPGVIASASISNFPDPVLDRLATVSPVVRIGKDAEVA

GLCAWLASDDAAYVTGADYAVNGGLYMG"
4181. .5212
/gene="cebr"
/note="synonym: XCC3356"
4181. .5212
/gene="cebr"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="AA042626.1"
/db_xref="GI:21114606"
/translation="MAKHPKRAATPAAPQWADIARMAGVSESTVSRLAGSPVVAERT RAYIKQAAAGYQVDFVARSRAKRSNTVSAVMMHALDQPLSDPFLMTMLLALAE ELTGYSMLISKLNRHQDWQVQLARSGSDGVIVLQSSSEHALDAAADGJPMAY WGSRIDQSYLISVGNDFQGGALATEHLIARSGRIAPLFGDDQLPEVAPRFAGYRQAL EOHGLEFTRHLARSHFLSDAYRLTKMLKKTDPGLFAASDVIAVGATRAIVEAG HRVPQDISLNGFDDIPLAAYSQPLTTVQDDGLAARLLVDRLLALTAGEAVIDSVEMP VKLVYRESA"
complement(5440. .6759)
/gene="sucI"
/note="synonym: XCC3357"
complement(5440. .6759)
/gene="sucI"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sugar transporter"
/protein_id="AA042627.1"
/db_xref="GI:21114607"
/translation="MSSTAPPLSPARLINALNAGFFVQVSPGLQOSNMSPITVNYLGAD HASTPYLWLAGPITGLVLOPVGALSRSVTRGCRMPYVVLGALVSLCICLAMPST ALMNAVCLLWLDANNVNEPRALNSVDLAPQRPGLYLTQSAFSLCILLAPPTP PLLVWQSDAANAHIPIYVITAAFIAGFSAASILLITARSVPFVPAAREIARNR KAGLGCATREBESALRDMPTWQAPVNLQWYALFQWYVILSLSTLFTLFTTE ANSHGFAGLVNGQIGFYNTIAFLAAFAVPPVRRVGPYTHAACLLAAGVGMVL PGIERWMLLPMIGIGLAWSMGNPYLMADSIPTPTGVYMGLEFLVLPMLIQ IVTLPLYEPLLLHGDPRNVIRLACALMLAAVAMLCVRIRKPKAAV"
7007. .9481
/gene="fyuA"
/note="synonym: XCC3358"
7007. .9481
/gene="fyuA"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="iron-dependent receptor"
/protein_id="AA042628.1"
/db_xref="GI:21114608"
/translation="MRERIAMSTLHRLHALACAVTTCCLAPLAAQDATPPAPATP PAADSAVNLDSVFTGTSTATTKLSSVSSTVGAEAIBQSAPRSTAEIFRNPIGIR SESSGCGNANAVRGIPVAGGAKFLQEDGIPUMFGDIAPGNADIPLRSDFTMD RAIKRGSGASTTSNAPGIIINFTDKGTGEGSVSGVSGRLDNDTRIDENYGPAPY EHWQIFGFPFQGDVDRDAGYTDKGGUKANILTRLFENGYVAVYVTRKILNDRAAGL PVTVSVRGDSGLPGGPFDFNDTLYSRNFRTPDVLGDNWQPRFTDLGDMHPIS RTTGAZAFDLGNGWLSDFRIADNSGRVSPFPPEVTDAAALSSIGGAGQOLVER GGANAGAYTGTALTRHFNVALNCLNFTVNDLSREFGDGKRTLNRMGYVTSROT LDMWTWNSYVQSLGRDRLLNVVDAAGVSRSONGLYVGTFFPMGDCCTITSYDVRVD VNPATVSLTFDSKLSIDGSLRVKMGDARGVSGTAIAQNLVNDGDIOPVEORVAT VDTANRPVYDNYLSLGGNYLINDDLGAPARVSRGARANADRLFLPVRIDGVSU TSDEAVNVVQTRAGLKKRRDGLSLFATAFQEQNFVTSORFNRSFKAGIEL EASYRYEGFTVNGVWTDIARDQITPTNGVPRQADFVWQLTFPSVRGDGYQFG VNLITGTEAYTQDSNOLKMPYQVNLFGDYRITDALTVALNVNLTFTGLTBAEEA TIPANGIIRARSAGRTTSLSLRYDF"
9515. .11428
/gene="XCC3359"
9515. .11428
/gene="XCC3359"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

```

/codon_start=1
/transl_table=11
/product="amylase or alpha amylase"
/protein_id="AA042629.1"
/db_xref="GI:21114609"
/translation="MIASPIDAALRASVAALDPSTAVATLARFDTHAPRLDLALS
TLVGHADVTTLPOWLSALGVVAOPALRDLDSRAPGFGQDMGLGSAYVDLR
ACTLRGVAERFVLOELGVYRLHLEPLFARAGDNDGGFANSDYQVPEFALGSDNLI
ALTARLRANISLCADFVLIHHTADHWAQARAGDTRYLDIYHFAURNAPDQDIT

```

Alignment Scores:

```

Pred. No.: 2,3e-86      Length: 11605
Score: 1116.50      Matches: 241
Percent Similarity: 56.51%      Conservative: 106
Best Local Similarity: 39.25%      Mismatches: 236
Query Match: 32.76%      Indels: 31
DB: 1      Gaps: 10

```

US-09-843-007A-2 (1-636) x AE012454 (1-11605)

```

QY 43 ArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsn 62
DB 9605 CGTTTCGATACGATCCCGCGCTTCGACGGCTGTCCACGCTCTACCGCGACCAT 9664
QY 63 GluAlaLeuLeuProMetLeuLeuMetLeuLeuAlaGlnAlaTrpGlnSerTyrSerGln 82
DB 9665 GCCGACTACGTCACCTGTTGCCACAGTGGCTGAGCGCACTGGCGTGGTGCACAGGCA 9724
QY 83 ArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrpIleLeu 102
DB 9725 CGCGACCGCGCTTCGGGCCTTCGACGACCTCGACGACACCGCGCA-----CCGGCGTTCGCG 9778
QY 103 SerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeuLysGly 122
DB 9779 CAGCAGGACATGCTGGCTTACAGCGCTACGCTAGTGGACCGTTCGCGGACGCTGGCGGT 9838
QY 123 LeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetPro 142
DB 9839 GTTCCGACGCGGTTCGATTCGAGGAACCTGGCGTGGCGTGGTGGTGGTGGTGGTGGT 9898
QY 143 LeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAsp 162
DB 9899 TTCCTGGCGCAGCGCGCGCGACACGACGACGCGGTTCGCTGAGGATACGCGCCAG 9958
QY 163 ValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGlu 182
DB 9959 GTGGAACCGCGCTTCGGCAGCAACGACCATCTGGTCGCACTCACCACCGCGCTACGCGC 10018
QY 183 AlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp 202
DB 10019 GCCAACATCAGCTGTGGCAGATTTCGGTTCGATCCACCAACCGCGCGACACCGCTTCG 10078
QY 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222
DB 10079 GCACAGCGCGCAGCGCTGGCGATACGCTTATCTCGACTACTACCATCCTTCGCGCAG 10138
QY 223 ArgArgMetProAsnGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisPro 242
DB 10139 CGCAACGACCATTCGACGACGACACCACTTCGCGTGGTGGTGGTGGTGGTGGTGGTGG 10198
QY 243 GlyGlyPheSerGlnLeuGluAspGlyArg---TrpValTrpThrPheAsnSerPhe 261
DB 10199 GGCAACTTCACCTGGTGGCAGCAGCGCGGAGTGGATGGACACCGCTTCATCCCTAT 10258
QY 262 GlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeu 281
DB 10259 CAGTGGGATTTGAATGGAGCAACCGCGGTGTTTCGGCGAGATGGCATTCGCGATGCTG 10318
QY 282 PheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLys 301
DB 10319 GAGCTGGCAATCTGGGTGGAGCGGTTCGACTCGATTCCACGCGCTACTCTGGGAAA 10378
QY 302 GlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsn 321

```

RESULT 13
AE005791/c

```

DB 10379 CGCCCGGCGCACCACTGTGTAACACGAGCGCGACACCATCTTGTGTGCATTGGGT 10438
QY 322 AlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisPro 341
DB 10439 GCGGTGGCGGACATCGTGGCGCTTCGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10498
QY 342 AspGlnValValGlnTyrIleGlyGln-----AspGluCysGlnIleGly 356
DB 10499 GCGGAGCTGCGCGCTACTTCGGCAGCGCGCTGTGACGCGCGGCGACGAGTGTCCAT 10558
QY 357 TyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeu 376
DB 10559 TATCACACAGCTGATCGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 10618
QY 377 LeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyr 396
DB 10619 CTGCAGGAGCTGATCGCGCACAGCCACCGCTGGCGCCAACTGGCGCTGGCTCAGTAC 10678
QY 397 ValArgSerHisAspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGly 416
DB 10679 GTGGCTGCCATGACGACATCGCTGGAACTGTGTGACGACGACGCGCGCGCTCGCTGCC 10738
QY 417 IleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGly 436
DB 10739 CGCAACCGCGCTTCGCTGCGGAGTGGCGAGTTCTACGCCAACCGCGTGGCGCGC 10798
QY 437 SerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGlyAsp-----CysArg 454
DB 10799 AGTTACGCGCGCGCGGAAAGTTTCAG-----AGCAGCGCGGATGGCGTGCATGGC 10849
QY 455 ValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHis----- 471
DB 10850 ACCAATGGCATGAGCGCGCATTTGGTGGCGTGGCGAGTGGCGACACCGCATGGC 10909
QY 472 -----AlaValAspArgIleLysLeuTyrSerIleAlaLeuSerThr 486
DB 10910 CGCGCGCGCGCGCTGGCGTGGATCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10969
QY 487 GlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspTrp 506
DB 10970 CCGGTGTTCGCTGATCTACATCGGCGATGAACTCGCGCTACCCACGACACCGCTTAC 11029
QY 507 SerGlnAspSerAsnLysSerAspSerArgTrpAlaHisArgProArgTyrAsnGlu 526
DB 11030 CTGACGATGCGACGACGCGCACGAAAGGACGCTGGTGTGTGTGTGTGTGTGTGTGTGT 11089
QY 527 AlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeu 546
DB 11090 GAGCTGGCGCGCGCGCGATGACCGCTCCACGTTGGCAGTACGGTGTACACGCGCTTG 11149
QY 547 ArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArg---LeuVal 565
DB 11150 CGTGATTTGATCGCGCTGGTGGCGCTGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 11209
QY 566 ThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsnAlaLeuLeuAla 585
DB 11210 AGCTCGCGCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11269
QY 586 PheGlyAsnPheSerGluTyrPro-----GlnThrValThrAlaHisThrLeu 601
DB 11270 GTGCACAATTCAGCGACGCTGCCACTCGCGTGGATCTCACCCAGACCGGACATGCCCT 11329
QY 602 GlnAlaMetProPheLysAlaHisAspLeuIleGly---GlyLysThrValSerLeuAsn 620
DB 11330 TGGCGCTGTG-----CTGCACACCGACGCGCATGGCGATGGCGCGCGGACCGCAT 11377
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGlu 634
DB 11378 ACGGAGCTGTGTCTTCGCGCTATGGCGTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGT 11419

```

LOCUS	AE005791	10822 bp	DNA	linear	BCT 12-JUN-2002
DEFINITION	Caulobacter crescentus CB15 section 117 of 359 of the complete genome.				
ACCESSION	AE005791	AE005673			
VERSION	AE005791.1	GI:13422441			
KEYWORDS					
SOURCE	Caulobacter crescentus CB15				
ORGANISM	Caulobacter crescentus CB15				
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales; Caulobacteraceae; Caulobacter.				
AUTHORS	1 (bases 1 to 10822) Nierman,W.C., Feldblyum,T.V., Leub,M.T., Paulsen,I.T., Nelson,K.E., Eissen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H., Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and Fraser,C.M.				
TITLE	Complete genome sequence of Caulobacter crescentus				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)				
MEDLINE	21173698				
PUBMED	11259647				
REFERENCE	2 (bases 1 to 10822) Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eissen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Leub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouri,H., Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and Fraser,C.M.				
AUTHORS	Direct Submission				
TITLE	Submitted (31-JAN-2001) The Institute for Genomic Research, 9712				
JOURNAL	Medical Center Dr., Rockville, MD 20850, USA				
FEATURES	Location/Qualifiers				
source	1. .10822 /organism="Caulobacter crescentus CB15" /mol_type="genomic DNA" /strain="CB15" /db_xref="taxon:190650" 141..452 /gene="CC1128" 141..452 /gene="CC1128" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK23112.1" /db_xref="GI:13422442" /translation="WKISITIAAGVAGSPFIAQALAGGSPEDQWACIQOAHNRDA AQWLECNADGGKLSDCVVDNSAAGKGFKAAMCIAEKLPMGAKTGTVKVPFRPGG A"				
gene	670..1176 /gene="CC1129"				
CDS	670..1176 /gene="CC1129" /note="identified by match to protein family HM"				
gene	/codon_start=1 /transl_table=11 /product="RNA polymerase sigma-70 factor, ECF subfamily" /protein_id="AAK23113.1" /db_xref="GI:13422443" /translation="MSDLASLRQRDLQYVRRTRDPATAEDLVQETFERLLAYGQ AKTVIRALGYGIALNLVRDHFRLAQRSAFQALDDI PCPPAPBQIIMHRRQYVEF SOVLDAWPLRRDVFIRRLHGHSSRQISKDLSLSAAVEKHVARALEQLREVARAE RAGGGRS"				
gene	1173..2096 /gene="CC1130"				
CDS	1173..2096 /gene="CC1130" /note="similar to SP:P23485 GB:M63115 PID:145931				

PID:537133	GB:U00096; identified by sequence similarity; putative"
/codon_start=1	
/transl_table=11	
/product="transcriptional regulator, putative"	
/protein_id="AAK23114.1"	
/db_xref="GI:13422444"	
/translation="MNRIAQEAALWAARLDADAQDPDAFWAKAADPANARAYEQVW RASQDPALSEALRYSEORFAAPRAAPVAPVAGLACGLAVFTAWPAQVMTVOQV IQTAGQORVVALADCTVILDGATRLDVLRTGRRQVELYRGEAFVDAHADARFFI VQAREGARVLTGADLRLGRGKLSVRRGKRLAPGTGVIIRRTALIELVGRQAFKRG FLSAVTRFDQADDMTGWLTDGVTGLARLVRNLNRASDKPTIIDPALGRQVARGF RLDDPEALARNALMHGFTIETSENGLELR"	
2172..5102	
/gene="CC1131"	
2172..5102	
/gene="CC1131"	
/note="similar to GP:9187833; identified by sequence similarity; putative"	
/codon_start=1	
/transl_table=11	
/product="TonB-dependent receptor"	
/protein_id="AAK23115.1"	
/db_xref="GI:13422445"	
/translation="WTFPRDIGDDMTAKTWTWLSASALAMKLACAPALAPAAKPTL AIPAGDLATALPYFSRATGLQVLADPALLKGRKTEGVGRQFEAREGLTELRGTGLNF VRNGGAAILRPMQORVSTAPVAAQAPVAPTAQAEPAEALVAVVGTGFANSLAR ALADKRAANVVDVAISADIKGTFPSNNIAEALQRPVGAITDRGEGGLFVRVGLGN FQVITLGRSAANVMNRDSCQSGQRFDTLPSELVAGVGVKSPGLASLDEGALGV VDVITFRDLIGRSTLALSAATSPQLAGTKDPRLSGLASWNTAEGTGVLAALVPDQ RTLQDRITGVGWSYADPASTLTKMTIAQGISLYPTAIRPTLEREDRERKGVLSAQT RPSDETEITVDASVTKLDHYDELTVSDQNVATLAGSAVTKGVLVCGTWNSTQITQI GREVDLADHNLVAGRIQKLEDWTLADIAAFAPSETNSNITRLKLGFGITTF SMTVDDVDAVPSVRLATADLGNPSLTPFRISMENNAITDDETALQFAERPLSGLL SRVQGVKRYERSDYNRADINTYTNLAGKFFGAEYFADFVSPVGLGVDGTLPTNWA MPDPAFWGLADKTQRADAKNSYQVSDIAAGYILGDSVEKVFNPALRGQLGVR VARTEQISAGHADNGSAIPVRYKSTYTVLPTANFALELTDTLQARFAAKVI TRPS LADLAPRLTNSGTVLTAVGNNPLKPFKAWQYDTELEWFAFGSALIAGAYKDIIT TFVYSQKTNIVDQZVYLLTSPNGSGAKIKGIELAYQHLFKRLPAPFDGLGFLANYT YTDTEATYPTLKDMVNVAKNSPLTAFYEKDRFAAKASYSWDDVLIQDVGKGLSA LNDSPAGSLDASVSVKITPTPTAAVAEQNLNTASOWQFVKGWFGFYTHYGRITLSFGL RAKF"	
5208..6737	
/gene="CC1132"	
5208..6737	
/gene="CC1132"	
/note="identified by Glimmer2; putative"	
/codon_start=1	
/transl_table=11	
/product="conserved hypothetical protein"	
/protein_id="AAK23116.1"	
/db_xref="GI:13422446"	
/translation="WLKLSLTHRRPCLSPVALVALCLAPVSLPAFAHGHAVQVGP SRPGRVAVKSERKQAGDHHVSHVSVLSKPGADPSQPPALIAGADVNEITTNARM AKHGLAWVATDHGPNHNSKLNDEAPPELLRSRVETPAVLQWGMELDTFAHSHS LILPSEERDALFISRYNRNREFPADPADQAEWALKEALRKALKKPEVVIANH PSRSATAVRGVLSVPSLBNWNTAPDVAIMEGAPGHQASLKPLKQPGDRGG YSRAPTMGGFDQMTAKLGFWDMLGGRRWITATSDSHRYHTDGGDFWPGYSKT YVKAHTYDDVLDGLRNGRVFTLGLVSEVETASVAGAKADIGRLEVAQSDVTY VIVRDPAKGNFSGSPSVARVDLIGELGPATDRSDANPTTTRVIRPAFTPSHARE GEMLSMTHLVKDVRRPVYIRVRCGTNGAEPLDPLGEDPWSDLFWANPFIITVASA R"	
complement (6759..7994)	
/gene="CC1133"	
complement (6759..7994)	
/gene="CC1133"	
/note="identified by match to protein family HM"	
/codon_start=1	
/transl_table=11	
/product="transporter, putative"	
/protein_id="AAK23117.1"	
/db_xref="GI:13422447"	
/translation="WTQTVSDASDRPRLIRILTYLMFMFPAMTTSVGVILPHVIKTF	


```

DREGEMARELVCHRYGVEQVQLWLSALDDASIRKALRLFLPGHETFNLYHSAGRS
RADWLIAGMNRNLTLLRGSGYQGVLPVGRVOTPTLRVLVDVDRSIANVPVPEAI
DVOLEHAGORENAQWRAPEDACDQGRCLQSLAQQAADIGHAGTARAVKATERYR
EAPLPDGLTGLQELCKEGLGAQETLIDIAOLYTHKLIITPRSDCGTGLFQSQHAE
APAILAORADASLAPOLYLPQRSDANDAKVSAHGLIIPATAASDPRLPAKH
KAVYTLIRARYLAQFPNEHYDKTQADFDCAGHALRAVGQIVPEFGWRRLAEALTPA
KGRAPPAQVLPALREGQCNVHGLQKDLWTQPPKPTFEGDLIRAMONVAKLPEALTPG
LQKQKBTIGTIGTRASIIQGLRLYLVKNGKALSATPAAFSLIDIVPRAIADPG
TTAIMQALDMQSGEMPLEEFVARGQSAWNGKLVRCSGWRMTISGPAAGAAPPWKX
RGGGSKATGSKATSPKXQARKKRAPI"
5414..6703
/locus_tag="pp4020"
5414..6703
/locus_tag="pp4020"
/notes="similar to GP:9951625; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, putative"
/protein_id="AA069614.1"
/db_xref="GI:24985720"
/translation="MFKQSAQVASYAQTPTNPLRPTLQGTQDADVLIIAGFSG
LATALRLTQAGRVTLLEASRVAMASGNGOALLGSCDMPPELQALGIERTOLW
ASMWAAEMRLPQRHGFIDYRQGLSLWTALPRVKMLBEALHDAHKKGYDALRL
IGRDELPONIDSPRYQALYDAKHAHLPKLAQGLASTIEAAGQIYEQSVLYSQ
ARDGYARTDRGEVRCQVLVLACNATIDRLDGLRRLLPVGSYQVATAPLDADFAHS
LUPRNSVDINDQFVDPYFRLTDPHRLLLFGGCTYLGSGIPKDVASATRYLERYFPOLA
GVAYDIAMGHHIDCSIKRTPDVGREGQRYLQGFSGHGYLPTLAARAVSDAILGDED
6376..7706
/locus_tag="PP4021"
6376..7706
/locus_tag="PP4021"
/notes="similar to GB:U07989, SP:P01596, SP:P01602,
SP:P01604, SP:P01834, PID:1335148, PID:1619960,
PID:185945, PID:185947, PID:185949, PID:468243,
PID:470602, PID:470608, PID:470612, PID:561696,
PID:561705, PID:563650, PID:619427, PID:619601,
PID:619602, PID:619613, PID:619615, PID:619616,
PID:632988, PID:642582, PID:663000, PID:663001,
PID:722576, PID:758085, PID:762937, PID:790811,
PID:860994, PID:896278, GB:U07989, SP:P01596, SP:P01602,
SP:P01604, SP:P01834, PID:1335148, PID:1619960,
PID:185945, PID:185947, PID:185949, PID:468243,
PID:470602, PID:470608, PID:470612, PID:561696,
PID:561705, PID:563650, PID:619427, PID:619601,
PID:619602, PID:619613, PID:619615, PID:619616,
PID:632988, PID:642582, PID:663000, PID:663001,
PID:722576, PID:758085, PID:762937, PID:790811,
PID:860994, and PID:896278; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="AA069615.1"
/db_xref="GI:24985721"
/translation="MSVYTTKDCQVIFVKDKGPRDAPVTHHHGKPLSADWDQMLF
FLAEGRYVYHDKRGHGSSQVWNGHDMHTADDOVAUVSLGTQGVAVHVGSTGGGE
VRYWARYPEUKKAXVLAIAVPLPVQGTGPKSPVFDGFOAQVANSRQAFQFP
VPTGFGYGNRPVGDASGIIQNNWRQMGISAKAHYDGIYAFSQDPTEDLKGIQP
VLVWEGDDQIVPYENSGLVSAKLLPNTGLTKLYEGYPHGMPHTHADVINADLLAFIOS
7826..8617
/locus_tag="PP4022"
7826..8617
/locus_tag="PP4022"
/notes="conserved domain protein"
/product="conserved domain protein"
/protein_id="AA069616.1"
/db_xref="GI:24985722"

```

```

/translation="MLIRAKYDEITQILNLVLTMLTKQTQPDPSAASLTNALRHFPAYS
CIGYXPQFLLATFNQYIDLFSGHYQLGSIYRTNTGLMRPHSPDSPPDGGGLN
AYACSGDFINWHDPSGHPVLKSKSFLPKTKTINORREALKAINENPVLBAFOGIA
ENSVTKSELKTLSDTKALNFTDLMEKNKLJSIDKQGLKYKLMDDFSLRQIILED
VMKNDKTLNLSLSAYPRANSPEPSLSSASLSPATNKLTRTKQGN"
8693..9337
/locus_tag="PP4023"
/notes="This region contains an authentic frame shift and
is not the result of a sequencing artifact; this gene is
interrupted by an IS element.; membrane protein, putative,
interruption-N; identified by match to PFAM protein family
HMM PF01925"
9347..11387
/insertion_seq="IS0024"
9432..9791
/locus_tag="PP4024"
9432..9791
Alignment Scores:
Pred. No.: 4,8e-35 Length: 304517
Score: 550.00 Matches: 164
Percent Similarity: 43.72% Conservative: 83
Best Local Similarity: 29.03% Mismatches: 228
Query Match: 16.14% Indels: 90
DB: 1 Gaps: 20
US-09-843-007A-2 (1-636) x AB016789 (1-304517)
QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
DB 56457 GCGCATTTTCGCGCGCTGATCAGCAAGCTGACTATATCGCGGAGTGGTGGTCAACACC 56398
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
DB 56397 TTGTGCTGTCGCGCTCTAT-----CCCTCGCCACCGCTGACGCGCTACGACATC 56344
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrLeuGlyAspLeuArgGluValIle 177
DB 56343 GCCGATACAAAGCGCGTGCACCTGACTACGCGGACATGCGCGGACGCGCGCTTTCATC 56284
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197
DB 56283 GCCAGGCCCAAGAGCGCGCTCGGGTGATCCCGGCTGGTTCATCACCAACACCCAGC 56224
QY 198 AsnGluHisGluTyrAlaGlnArg-----CysAlaAlaGlyAspProLeuPheAsp 214
DB 56223 GACCAGCACCGCTGCTTCAGCGCGCGCGCCGCCAACCAAGCGCGCAGCAAGCGCGCGAG 56164
QY 215 AsnPheTyrTyrIlePheProAspArgArgMetProAspGln---TyrAspArgThrLeu 233
DB 56163 -----TTCACGCTGTGTGCGAC-----GACGACCAAGAGTACGACGCGACG--- 56122
QY 234 ArgGluIlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArg 252
DB 56121 CGCATCATCTTCCTCGACACCGAGAGAGTCCCACTGGGACCTGGGACCGCTCGCGGCGAG 56062
QY 253 TrpValTyrThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsnProTyrVal 272
DB 56061 TACTTCTGGCACCGGTTCTACTCCACGACGCGGACCTCACTTCGACAAACCCGCGGTG 56002
QY 273 PheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArg 292
DB 56001 CTCAGGCGGTGATCGCGCTCATCGGCTTCTGGCTGACCTGGGTGTAGATGGCTTCGGG 55942
QY 293 MetAspAlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGln 312
DB 55941 CTGACGCGGATCCCTTACCTGATCGAGCGGCGGACCAACCAACCAACCAACCTCGCGAA 55882
QY 313 AlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPheValPhe 332
DB 55881 ACCCAGCATGTCTCAGGCGGATCGCGCGGAGATCGACGCCCACTACCCCGCATCGCATG 55822

```

```

QY 333 PheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGln----- 350
Db 5821 CTGCTGGCGGAGCAACAGCTGGCTGAAGACACCGCGCCGCTACTTCTGGCGAAGCGGCAC 55762
QY 351 ---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeu 369
Db 55761 GCGGACGATGCCATGCTTCCACITTCCTGGTGGATCGCGCATGTACATGCGCTTG 55702
QY 370 AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389
Db 55701 GGCATGGAAGACCGCTTCCCTATCCGACATCTCGCGCAGACCCGCGAAATCCCGGCC 55642
QY 390 HisThrAlaTyrValAsnTyrValArgSerHisAspIleGlyTyrThrPhe---Ala 408
Db 55641 AACTGCCAGTGGCGATCTTTTGGCAACCATGATGACACTACCTGGAAATGCTCAC 55582
QY 409 AspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArg 428
Db 55581 GACCGGAGCGGACTACCTG----- 55561
QY 429 PhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnPro 448
Db 55560 -----TGGAACTACTACGCC 55546
QY 449 SerThrGlyAspCysArgValSer---GlyThrAlaAlaAlaLeuValGlyLeuAlaGln 467
Db 55545 GAAGACCGCGCGCGCATCAACTGCGCATTCGCGCGCGCTCGCGCCCTGCTGCAG 55486
QY 468 AspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeuSerThrGly 487
Db 55485 COTGACCGCTCGG-----CGCATAGACGTGCTGACCGACGCTGCTGTGTCGATGCC 55435
QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
Db 55434 GCACGCCAACCTGTATTACGGCGACGACATAGCATGGCGGACAAATATCATCTGGT 55375
QY 503 AspAspAsp-----TrpSerGlnAspSerAsn----- 511
Db 55374 GACCGCGATGGGCTTGGCACACCCATCGAGTGGTGGCGACACCGCAACGCGCGTTCCT 55315
QY 512 LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeuTyr----- 529
Db 55314 CCGCGCGACCGCGCGCTGGTG---CTGCGCGCGATCATGACCGCGCTCTACGGTTAC 55258
QY 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyr 543
Db 55257 CAGACTGTCAACGTCGAGCGCAAGCCACCGACCGCATTCATCGCTATATCGCGAGTATACCG 55210
QY 544 GlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyArg 563
Db 55209 AACTGACCCCGCGCATGCTGGGTGCGGACAGCAGCAAAAGCGCTTGGCGCGCGCAGC 55150
QY 564 LeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg----- 579
Db 55149 CTGCGTACCTCAGCGCCAGCAACCGCGCATCTTGCCTATATCGCGAGTATACCGAC 55090
QY 580 -----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThr 595
Db 55089 GCGGAGGCAATACCGAGTGTATCTCTCGGTGGCCAAAGTGTCTCGCGCCCGCCAGCGCT 55030
QY 596 ValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyLys 615
Db 55029 GCGGAACTCGAATGTGCATATACCGCGCAAGTACCGGTGAGATGCTGGGTGGCAGC 54970
QY 616 ThrVal-----SerLeuAsnGlnAspLeuThrLeuGlnProTyrGlnValMet 631
Db 54969 GCGTTCGCCCAATAGCGCAACTGCGGTTCTGCTGACCTTCCACCTTATCGCTTCTAC 54910
QY 632 TrpLeuGluIleAla 636
Db 54909 TGGTTCCTGCTGGCC 54855

```

RESULT 15

```

AL646077/c
LOCUS
DEFINITION
Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
segment 2/11.
ACCESSION
AL646077 AL646053
AL646077.1 GI:17430642
KEYWORDS
SOURCE
ORGANISM
Ralstonia solanacearum
Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE
1
AUTHORS
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
Sguier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE
Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL
Nature 415 (5871), 497-502 (2002)
MEDLINE
21681879
PUBMED
11823852
REFERENCE
2 (bases 1 to 189050)
AUTHORS
Boucher,C.A.
TITLE
Direct Submission
JOURNAL
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dedu, 75010 Paris, France, IMGM CNRS
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
COMMENT
http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
Location/Qualifiers
1..189050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
79..1500
/notes="RS49+ or RS06146
LONG DIRECT REPEAT 49 +
predicted by Homology"
/evidence=not_experimental
83..2971
/gene="RSP0175"
/notes="synonym: RS04687"
83..2971
/gene="RSP0175"
/functions="miscellaneous; hypothetical/global homology"
/notes="product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE VGR-RELATED PROTEIN"
/protein_id="CAD17326.1"
/db_xref="GI:17430643"
/db_xref="GOA:Q8XID8"
/db_xref="SPTREMBL:Q8XID8"
/translaton="MDFKTLQHLFAPAHRLYALEGEGPTAEALVAEWLGRALSELFL
EMRVAASANARFALESFIGORVLTVTTLADGTQARTGLRQAEQAGDSGLARYL
TVVFWLATQQRSHQVFNRPRLDILEVLSVPYAAVRFPAAGADRMFAFTRSH
IAQRETDVRFVTFLLAAGLGTTFVEQAPSGHALLIFADSPRIADTASRAEGGI
RYRPAHSGERDAIQALICHSRTAVGGVAVAAWDAQAKRAFHVPFRFCGIAGSPDA

```


Qy	158	Ser	Ser	Tyr	Arg	Asp	Val	Asn	Pro	Ala	Leu	Gly	Thr	Ile	Gly	Asp	Leu	Arg	Glu	Val	Ile	177		
Db	92764	GCC	GAA	TAA	CCG	CGG	CGT	GCA	CCC	GCA	CTACT	ACG	CGC	ACG	CGC	CGC	CGC	CGC	CGC	CGC	CGC	92705		
Qy	178	Ala	Ala	Leu	His	Glu	Ala	Gly	Ile	Ser	Ala	Val	Val	Asp	Phe	Ile	Phe	Asn	His	Thr	Ser	197		
Db	92704	GCC	GAG	GCG	GCA	CGG	CGC	GCT	CGC	GCT	AT	TAC	GAG	CTG	GT	CAT	CAC	ACC	ACT	TCG		92645		
Qy	198	Asn	Glu	His	Glu	Trp	-----	Ala	Gln	Arg	Cys	Ala	Ala	Gly	Asp	Pro	Leu	Phe	Asp		214			
Db	92644	GAC	CAC	ACG	CCG	TGG	TTT	CAG	CGC	GCG	CGC	CGC	CGC	CGC	CGC	CGC	CGC	CGC	CGC	CGC	92585			
Qy	215	Asn	Phe	Tyr	Tyr	Ile	Phe	Pro	Asp	Arg	Arg	Met	Pro	Asp	Gln	Tyr	Asp	Arg	Thr	Leu	Arg	234		
Db	92584	-----	TT	CAC	G	TGG	TCG	GAC	CCAC	-----	GAC	AGAA	TAT	CGC	CGC	CGC	CGC	CGC	CGC	CGC	92540			
Qy	235	Glu	Ile	Phe	Pro	Asp	Gln	His	Pro	Gly	Phe	Ser	-----	Gln	Leu	Glu	Asp	Gly	Arg		252			
Db	92539	AT	CAT	CTT	CAT	GCAC	ACG	CGC	CT	CAA	CT	GGAC	CTG	GAC	CCG	CGT	GTG	CGC	GAC	GCC	--- 92483			
Qy	253	Trp	Val	Trp	Thr	Thr	Phe	Asn	Ser	Phe	Gln	Trp	Asp	Leu	Asn	Tyr	Ser	Asn	Pro	Trp	Val	272		
Db	92482	TAC	TACT	GGC	ACG	CGT	CTT	CTAT	TGC	ACC	AGC	CGG	AC	CT	GA	CT	TCG	AC	ACC	CCG	CGT	92423		
Qy	273	Phe	Arg	Ala	Met	Ala	Gly	Glu	Met	Leu	Phe	Leu	Ala	Asn	Leu	Gly	Val	Asp	Ile	Leu	Arg	292		
Db	92422	CTC	ACG	CGC	CGT	GTG	CGG	CGT	GAT	GAA	TTT	CTG	CT	CA	ACT	GGC	GTG	TGA	CGC	CGC	CGT	92363		
Qy	293	Met	Asp	Ala	Val	Ala	Phe	Ile	Trp	Ile	Trp	Ile	Ser	Gln	Met	Gly	Thr	Ser	Cys	Glu	Asn	Leu	Pro	312
Db	92362	CTC	GAT	CGC	GTG	CGC	CTT	ACT	GTG	GAG	CGG	AGC	CGC	CGC	CA	CGA	AG	AA	CTT	CGG	CA	CGC	92303	
Qy	313	Ala	His	Ala	Leu	Ile	Arg	Ala	Phe	Asn	Ala	Val	Met	Arg	Ile	Ala	Ala	Pro	Ala	Val	Phe	332		
Db	92302	ACG	CAC	CGC	GTG	CTG	CGC	AAG	ATC	CGC	GCG	CGAT	GGA	CGC	GCG	GA	TTT	CA	AAG	ACCG	CGCTG 92243			
Qy	333	Phe	Lys	Ser	Glu	Ala	Val	Ala	Phe	Ile	Trp	Ile	Trp	Ile	Trp	Val	Gln	Tyr	Ile	Gly	Leu	--- Asp 351		
Db	92242	CTC	CTG	CGC	GAA	CC	CAAC	CAG	TGG	CCG	CCG	AAG	A	CAC	AC	AGAA	TAT	CTT	CGG	CA	CGC	92183		
Qy	352	Glu	Cys	Gln	Ile	Gly	Tyr	Asn	Pro	Leu	Gln	Met	Ala	Leu	Leu	Trp	Asn	Thr	Leu	Ala	Thr	371		
Db	92182	GAG	TGCC	ACAT	GCC	CTT	CC	ACT	TTT	CGC	TGAT	GCC	CGC	GAT	TAC	AT	CGCAT	CGC	GAT	CGC	CGG 92123			
Qy	372	Arg	Glu	Val	Asn	Leu	Leu	His	Gln	Ala	Leu	Thr	Tyr	Tyr	Arg	His	Asn	Leu	Pro	Glu	His	Thr	391	
Db	92122	GAG	GAC	CGC	TTT	CCG	GAT	ACC	CGAC	AT	CAT	GCG	CGC	GAC	AGC	CCG	CGG	GTG	CGC	CC	CC	92063		
Qy	392	Ala	Trp	Val	Asn	Tyr	Val	Arg	Ser	His	Asp	Asp	Ile	Gly	Tyr	Thr	Phe	-----	Ala	Asp	Glu	410		
Db	92062	CAG	TGG	CGC	AT	CTT	CTC	TGG	CAAC	AC	CA	CGAC	GAG	CT	GAC	CTG	GAG							

Qy	505	Asp	-----TrrpSerGlnAspSerAsn-----	-----LysSer	513
Db	91795	GAGGGCGTGGCAGCGCCGATGAGTGGTCCGCCGACCGCAACGGCGGCTTCTGGCGCCG			91736
Qy	514	AspAsp-SerArgTrrpAlaHisArgProArgTyr-----	-----AsnGluAla	527	
Db	91735	GATCCGGAGGAGCTGTGTCTGCCGCCATCATGGCTCGCTGTACGGCTATGAATCGTTC			91676
Qy	527	LeuTyr--AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnLeuTyrGlnGlyLeu	546		
Db	91675	AACGTGGAGCGCAGACGGCGCAC-----	-----GGCCATTCGCTGTGAATCTGGACG	91628	
Qy	547	ArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThr	566		
Db	91627	CGCGCGCTGTGGTCCACGGCGAAGCGCCACCGCGTGTGGCGCGCGAGATCCAGTTC	91568		
Qy	567	PheAsnThrAsnAsnLysHisIleLeuGlyTyrIleArg-----	-----AsnAsnAla	582	
Db	91567	CTCAGCGCGCCAAACCGCAAGGTGTGGCCTATGCGCGCGCTGGAAGCGGAGGCCGCC	91508		
Qy	583	LeuLeuAlaPheClyAsnPheserGluTyrProGlnThrVal-----	-----ThrAlaHis	599	
Db	91507	ATCCTGTGGGTGGCGAATCTGTCCCGGCGCTCGCAGCGCGTGGAGCTGGAACCTGGCCGCG	91448		
Qy	600	ThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyLysThrVal-----	617		
Db	91447	TATCGCAGCGCGTGGCGTG-----	-----GAGTTGATCGCGGAGACGGCCTTCCCGCG	91397	
Qy	618	-----SerLeuAsnGlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIle	635		
Db	91396	ATCGGCAGTGCCTATCTGTGTGACGTCCGCGCCTATGCCTTCTACTGGCTGGAGTGC	91337		

Search completed: November 8, 2003, 21:58:11
Job time : 5786 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2003, 17:31:29 ; Search time 413 Seconds
(without alignments)
4157.004 Million cell updates/sec

Title: US-09-843-007A-2

Perfect score: 3408

Sequence: 1 MLPTQGVGLIQLVLRIL.....VSLNQDLTLQYQVWMLA 636

Scoring table: BIOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgm2.1/USPTO.spool_p/US09843007/runat_07112003_170419_25336/app_query.fasta_1.775
-DB=N_Geneseq_19Jun03 -QWTF=fastcap -SUFFX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNIT5-bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MTN=0 -ALIGN=15
-MODE=LOCAL -OUTFT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09843007 -CGEN 1 1 490 @runat_07112003_170419_25336 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DSLOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03.*
1: /SIDSI/cgdata/geneseq/geneseq-embl/NA1980.DAT.*
2: /SIDSI/cgdata/geneseq/geneseq-embl/NA1981.DAT.*
3: /SIDSI/cgdata/geneseq/geneseq-embl/NA1982.DAT.*
4: /SIDSI/cgdata/geneseq/geneseq-embl/NA1983.DAT.*
5: /SIDSI/cgdata/geneseq/geneseq-embl/NA1984.DAT.*
6: /SIDSI/cgdata/geneseq/geneseq-embl/NA1985.DAT.*
7: /SIDSI/cgdata/geneseq/geneseq-embl/NA1986.DAT.*
8: /SIDSI/cgdata/geneseq/geneseq-embl/NA1987.DAT.*
9: /SIDSI/cgdata/geneseq/geneseq-embl/NA1988.DAT.*
10: /SIDSI/cgdata/geneseq/geneseq-embl/NA1989.DAT.*
11: /SIDSI/cgdata/geneseq/geneseq-embl/NA1990.DAT.*
12: /SIDSI/cgdata/geneseq/geneseq-embl/NA1991.DAT.*
13: /SIDSI/cgdata/geneseq/geneseq-embl/NA1992.DAT.*
14: /SIDSI/cgdata/geneseq/geneseq-embl/NA1993.DAT.*
15: /SIDSI/cgdata/geneseq/geneseq-embl/NA1994.DAT.*
16: /SIDSI/cgdata/geneseq/geneseq-embl/NA1995.DAT.*
17: /SIDSI/cgdata/geneseq/geneseq-embl/NA1996.DAT.*
18: /SIDSI/cgdata/geneseq/geneseq-embl/NA1997.DAT.*
19: /SIDSI/cgdata/geneseq/geneseq-embl/NA1998.DAT.*
20: /SIDSI/cgdata/geneseq/geneseq-embl/NA1999.DAT.*
21: /SIDSI/cgdata/geneseq/geneseq-embl/NA2000.DAT.*
22: /SIDSI/cgdata/geneseq/geneseq-embl/NA2001A.DAT.*
23: /SIDSI/cgdata/geneseq/geneseq-embl/NA2001B.DAT.*
24: /SIDSI/cgdata/geneseq/geneseq-embl/NA2002.DAT.*
25: /SIDSI/cgdata/geneseq/geneseq-embl/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3408	100.0	1939	22	AAF61709	Amylosucrase PCR d
2	3408	100.0	2914	21	AAAI1732	N. polysacchara a
3	3408	100.0	4173	21	AAA07380	Neisseria polysacc
4	3408	100.0	6878	22	AAF61711	Expression vector
5	3360	98.6	1910	22	AAF61710	Amylosucrase PCR d
6	3360	98.6	6851	22	AAF61712	Expression vector
7	3141.5	92.2	2883	17	AAOT09860	Neisseria polysacc
8	3141.5	92.2	2883	17	AAI11179	Neisseria polysacc
C 9	1164	34.2	29559	23	AA559346	Propionibacterium
10	560.5	16.4	22934	23	AA559613	Propionibacterium
11	541.5	15.9	3303	23	AA554127	Pseudomonas aerugi
12	529.5	15.5	1794	22	AAH67492	C glutamicum codin
13	529.5	15.5	1981	24	AB565343	DNA encoding C. gl
14	529.5	15.5	34980	22	AAH68531	C glutamicum codin
15	529.5	15.5	34980	22	AAH68532	C glutamicum codin
16	509.5	15.0	4403765	22	AAI99683	Mycobacterium tube
17	509.5	15.0	4411529	22	AAI99682	Mycobacterium tube
18	504	14.8	1704	17	AAI11413	Sequence encoding
19	488	14.3	3600	17	AAI13227	Thermotable enzym
20	400.5	11.8	1136	23	AA559731	Propionibacterium
C 21	398.5	11.7	1113	23	AA559702	Propionibacterium
22	393	11.5	1704	24	AEQ67835	Listeria innocua D
23	393	11.5	684707	24	AEQ67196	Listeria innocua D
C 24	393	11.5	3011208	24	AEQ69245	Listeria innocua D
25	375.5	11.0	2944528	24	ABA03041	Listeria monocytog
26	373	10.9	1683	24	AEQ68051	Listeria monocytog
27	373	10.9	1683	24	ABQ70041	Listeria monocytog
28	364	10.7	1102	22	AAE71764	Corynebacterium gl
C 29	362.5	10.6	29912	23	AA559507	Propionibacterium
C 30	349.5	10.3	2944528	24	ABA03041	Listeria monocytog
31	339	9.9	1755	23	ABL26275	Drosophila melanog
32	333.5	9.8	1636	24	ABN66288	Streptococcus poly
33	332.5	9.8	1790	23	ABL26277	Drosophila melanog
34	331	9.7	1620	25	ABX08486	DNA encoding alpha
35	328.5	9.6	1650	23	AA554857	Staphylococcus aur
36	328.5	9.6	1704	23	ABU14333	Drosophila melanog
C 37	326	9.6	16592	18	AAV74364	Staphylococcus aur
38	324	9.5	4128	23	ABL26276	Drosophila melanog
39	323.5	9.5	1638	23	AA551873	Staphylococcus aur
40	319.5	9.4	1623	24	ABN66287	Streptococcus poly
C 41	319.5	9.4	2155561	24	ABN71527	Streptococcus poly
42	318.5	9.3	1163020	24	ABQ67197	Listeria innocua c
43	317	9.3	1653	25	ABT14986	Pathogen specific
44	317	9.3	1656	22	AAH53337	S. epidermidis ope
45	317	9.3	1683	24	ABN90853	Staphylococcus epi

ALIGNMENTS

RESULT 1
AAF61709
ID AAF61709 standard; DNA; 1939 BP.
XX
AC AAF61709;
XX
DT 12-JUL-2001 (first entry)
XX
DE Amylosucrase PCR derived DNA fragment SEQ ID 4.
XX
KW Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
NM poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
XX
OS Unidentified.
XX
FN WO200125449-A2.
XX

PD 12-APR-2001.
 XX
 PF 04-OCT-2000; 2000WO-EP09695.
 XX
 PR 07-OCT-1999; 99DE-1048408.
 XX
 PA (AXIV-) AXIVA GMBH.
 XX
 PI Bengs H, Polakowski T, Held A, Gallert K;
 XX WPI; 2001-328330/34.
 DR
 XX Amylosucrase immobilized as fusion protein with anchoring sequence,
 PT useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
 PT
 XX
 PS Claim 3; Page 27-28; 38pp; German.
 XX
 CC This invention describes a novel amylosucrase (AS), immobilizable on a
 CC solid phase, which comprises a fusion protein (FP) of functional units
 CC of AS, an anchoring sequence, and optionally additional auxiliary
 CC sequences. The invention also describes (1) nucleic acid (I) encoding
 CC FP; (2) expression vector containing (I) and able to express FP in a
 CC host cell; Escherichia coli containing the vector of (2); (3) anchoring
 CC sequence, or its functional variants or fragments, of at least 8
 CC nucleotides that encodes an epitope, a high-affinity binding partner or
 CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS
 CC comprising glutathione-Sepharose; (5) combined, stable catalyst (A)
 CC comprising FP immobilized on Sepharose for production of (II).
 CC AS is used in production of poly(1,4-alpha-glucans) (II), useful for
 CC producing films, as food additives, as starting materials for
 CC cyclodextrins and as auxiliaries in pharmaceutical formulations.
 CC Immobilized AS makes possible efficient, inexpensive and continuous
 CC production of poly(1,4-alpha-glucans) (II), and it can be used
 CC repeatedly. Compared with known methods, specificity is improved
 CC (increased yield of (II) and reduced formation of palatinose) and
 CC reaction is complete within 24 hours, compared to 48-72 hours for batch
 CC methods. This sequence encodes a PCR derived amylosucrase (EC 2.4.1.4)
 CC fragment described in the invention.
 XX
 SQ Sequence 1939 BP; 453 A; 605 C; 482 G; 395 T; 4 other;
 Alignment Scores:
 Pred. No.: 0 Length: 1939
 Score: 3408.00 Matches: 636
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0
 US-09-843-007A-2 (1-636) x AAF61709 (1-1939)
 QY 1 MetLeuThrProThrGlnGlnValGlyLeuLeuLeuGlnTyrLeuLysThrArgIleLeu 20
 DB 12 AUGTTGACCCCGCCGACGAGTCGGTTGATTTTACAGTACCTCAAAACACGCGATCTTG 71
 QY 21 AspIleTyrThrProGlnGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe 40
 DB 72 GACATCTACACGCCCGCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 131
 QY 41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
 DB 132 TCGCGCGCGATGATACGATTCCTCCCAACTGATGACGAACTCGACAGCGGTGACGGC 191
 QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr 80
 DB 192 AACAAAGAACCCCTGCTGCTATGCGGAAATGCTGCTGCGCGGCGGCGGCGGCGGCGG 251
 QY 81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrp 100
 DB 252 TCCCAAGCGCACTCATCTTTAAAGATATCGATATCGCGCGCGGCGGCGGCGGCGGCGG 311

QY 101 IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
 DB 312 AFTTTGTCCAAACAAGTCGGCGGCGGTGTGCTACGTTGATTTGTTGCGCGGCGATTG 371
 QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGlnLeuGlyLeuThrTyrIleHisLeu 140
 DB 372 AAGGCGTTGAAGATAAAATTTCTTATTTCAAGAGCTTGGTTTGACTTATCTCGACCTG 431
 QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
 DB 432 ATGCGCGTGTAAATGCTCCCTGAAGGCAAAAGCGGCGGCTATGCGGTGACGAGCTAC 491
 QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
 DB 492 CGCGATGTCAATCCGCGCACTGGGCACAAATAGGCGCACTTGGCGAAGTCATTGCTGCGCG 551
 QY 181 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200
 DB 552 CACGAGCGCGCATTTCCGCGGTGCTGATTTATCTTCAACCAACACCTCCACGACACAC 611
 QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
 DB 612 GAATGGGCGCAACGCTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 671
 QY 221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
 DB 672 CCCGACCGCGGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 731
 QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsnSer 260
 DB 732 CACCGCGCGCGCTTCTCGCAACTGGAAGACGAGCGTGGGTGTGCGACGACCTTCAATTCC 791
 QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280
 DB 792 TTCCAATGGGACTTGGAATTACAGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 851
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300
 DB 852 CTGTTCTTCTGCAACTTGGGCGGTGACATCTGCGTATGATGCGGTGCGGTTTATTAGG 911
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
 DB 912 AACCAAAATGGGACCAAGCTGCGAAACCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 971
 QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
 DB 972 AATGCGGTATGCGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1031
 QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
 DB 1032 CCGGACCAAGTCGTCCAATACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1091
 QY 361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380
 DB 1092 CAATGGGCGGTGTTGTGGACACCTTGGCCAGCGGCGGCGGCGGCGGCGGCGGCGGCG 1151
 QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400
 DB 1152 CTGACCTACCGCCCAACCTCGCGGAGCATACCGCGTGGGTCAACTACGTCGCGGCGGCG 1211
 QY 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420
 DB 1212 GAAGCATCTGGCTGACGCTTTTGGCGATGAGACCGCGGCGGCGGCGGCGGCGGCGG 1271
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
 DB 1272 GACCAACCGCAATCTCTCAACCGCTTCTTCTGTCACACCGTTTTCGACGCGGCGGCG 1331
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
 DB 1332 GCGGTACGTTCCATACACCCAGCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1391
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480

Db 1392 GCATTGGTGGCTGGCGACAGACGATCCCGCGCTGACCGCATCAACTTGTAC 1451
 Qy 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
 Db 1452 ACATTCGCTTTCAGTACCGCGCGCTGCGCGCTGATTTACCTAGGCGACGAAGTGGGTACG 1511
 Qy 501 LeuAsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis 520
 Db 1512 CTCATATGACGACGACTGGTGGCAGACAGCAATAGAGCGACGACGCGGTGGGGGCAC 1571
 Qy 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
 Db 1572 COTCCGCGCTCAACAGAACCTGTACGCGCAACCAACGATCCGTGACCGCGACCGCGG 1631
 Qy 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 Db 1632 CAATATCATCGGCTTGGCCATATGATTCGCTCGCGCAAGCAATCGCGCTTCGAC 1691
 Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
 Db 1692 GCGCGCAGCGCTGTATCATTTCAACACCAACCAACGATCATCGGCTACATCGCAAC 1751
 Qy 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
 Db 1752 AATGGCTTTTGGCATTCGTAACCTCAGCGAATATCCGAACCTTACCGCGCATACC 1811
 Qy 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyThrValSerLeuAsn 620
 Db 1812 CTGCAAGCCATGCCCTTCAAGCGCAGCACTCATCGGTGGCAAACTGTGAGCTGAAT 1871
 Qy 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla 636
 Db 1872 CAGGATTTGACCTTCAGCCCTATCAGGTGATGCGGTGCGAATCGCC 1919

RESULT 2

AAAL1732

ID AAAL1732 standard; DNA; 2914 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX New nucleic acid encoding a branching enzyme, useful for in vitro
 PT synthesis of branched glucans and to prepare transgenic plants
 PT producing modified starch
 XX
 PS Disclosure; Page 99-102; 115pp; German.
 XX

CC This invention describes a novel nucleic acid (I) isolated from
 CC *Neisseria* which encodes a branching enzyme (II). (I) is used for
 CC recombinant production of (II) subsequently used in the in vitro
 CC production of alpha-1,6-branched alpha-1,4-glucans. It is also used to
 CC prepare transgenic plants that produce starches with modified properties.
 CC (III) are used as binders for tablets, carriers for pharmaceuticals,
 CC flavors and perfumes and powdered additives, packaging materials,
 CC ultra-violet light adsorbers in sunscreens and also for any of the usual
 CC applications of starch in foods, papermaking, as textile size, in soil
 CC stabilization, as wetting agent for agricultural chemicals, as polymer
 CC additives etc. Fragments of (I) are useful as PCR primers and antisense
 CC molecules or ribozymes for inhibiting expression of (I), and the
 CC regulatory region of (II) can be used to control expression of
 CC heterologous sequences in host cells. (I) provides an inexpensive method
 CC for producing alpha-1,6-branched alpha-1,4-glucans (III), producing
 CC products that can be tailored for particular applications, particularly
 CC by controlling the degree of branching. Starch from transgenic plants
 CC has increased gel strength; reduced phosphate content; reduced peak
 CC viscosity; lower pasting temperature and granule size and/or altered
 CC sidechain distribution. This sequence encodes an amylosucrase isolated
 CC from *Neisseria polysaccharia* which is described in the method of the
 CC invention.

SQ Sequence 2914 BP; 662 A; 868 C; 773 G; 611 T; 0 other;

Alignment Scores:

Pred. NO.: 0 Length: 2914
 Score: 3408.00 Matches: 636
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-843-007A-2 (1-636) x AAAL1732 (1-2914)

Qy 1 MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20
 Db 957 ATGTTGACCCCGACGACGAGTCGGTTTGTATTTACAGTACCTCAAAACACGATCTTG 1016
 Qy 21 AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe 40
 Db 1017 GACATCTACACGCGCGCAACAGCGCGCGCATCGAAAAATCCGAAGACTCGCGCGCAGTTT 1076
 Qy 41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
 Db 1077 TCGCGCGCATGGATGACGCAATTTCCCAAACTGATGACGACTCGACAGGTGTACGGC 1136
 Qy 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTyrGlnSerTyr 80
 Db 1137 AACAAAGAACCTGCTGCTGCTATGCTGGAATGCTGCTGCGCGAGGCGATGCGAAAGCTAT 1196
 Qy 81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrp 100
 Db 1197 TCCCAACGCAACTCATCTCTTAAAGATATCGATTCGCGCGCGGAAACCAACCCCGATTCG 1256
 Qy 101 IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
 Db 1257 ATTTTGTCCAAACAAACAAAGTCCGCGCGGTGTGCTGCTGATTTGTTTCCGCGGATTTG 1316
 Qy 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
 Db 1317 AAGGGCTTGAAGATATAAATTCCTTTATTTTCAAGAGCTTGGTTTGTACTTATCTGACCTG 1376
 Qy 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
 Db 1377 ATGCGCGCTGTTTAAATGCGCTGAGGCGAAAGGACGCGGCGCTATGCGGTACAGCTAC 1436

(PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
 (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Buettcher V, Quanz M;

WPI; 2000-317992/27.

P-PSDB; RAW90979.

QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu 180
 DB 1437 CGCGATGCAATCGGCACATGGGCACATAGGACATTTGGCGAGATCATTTGCTGGCTG 1496
 QY 181 HisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSerAsnGluHis 200
 DB 1497 CAGAGCGCGCATTTCCGCGCTGCTGATTTTATCTCAACACACCTCCACAGAACAC 1556
 QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
 DB 1557 GAATGGCGCAACGCTCGCGCGCGCGCGCGCTTTTCGACAAATTTCTACTATATTTTC 1616
 QY 221 ProAspArgMetProAspGlnTyrAspArgThrIleuArgGluIlePheProAspGln 240
 DB 1617 CCGACCGCGCGATGCGCGACCAATACGACCGCACCTCGCGCAATCTTCGCCGACCAAG 1676
 QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTrpThrPheAsnSer 260
 DB 1677 CACCGCGCGCGCTTCTCGCACTTGAAGACGACGCTGGGTGTGGACGACCTTCAATTC 1736
 QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280
 DB 1737 TTCGAATGGGACATGAATACAGCAACCCGCTGGGTATTCGCGCAATGGCGGCGAAATG 1796
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300
 DB 1797 CTGTTCCTTCCCACTTGGCGGTGACATCTCTCGTATGATCGGTGCTTTATTGG 1856
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
 DB 1857 AAACAAATGGGACAGCTCGCAAACTTCGCGACGCGCACCGCTCATCTCGCGGTTC 1916
 QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
 DB 1917 AATGCCGTATGCTGTATGCGCGCGCGCGCTTCTTCAATCCGAAAGCAATCGTCCAC 1976
 QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
 DB 1977 CCGACCAATCTGTCCTCAATATCTGGGACGAGAGCAATGCAATCGTTACACCCCTCG 2036
 QY 361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAla 380
 DB 2037 CAAATGGCATTTGTGGAAACACCTTGCACGCGGCAAGTCAACCTGCTCCATCAGGCG 2096
 QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400
 DB 2097 CTGACCTACCGCCACACACCTTCCCGAGCATACCGCTTGGGTCAACTACGTCGCGACCC 2156
 QY 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420
 DB 2157 GAGCAGATCGCTGGACGTTTGGCGATGAGACGCGCATATCTGGGCATAAGCGGCTAC 2216
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
 DB 2217 GACACCGCAATCTCTCAACCGCTTCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2276
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
 DB 2277 GCGGTACCGTTCCATACACCCACGACGACGACGCGCATGCGGTGTCAGTGTGACGCGCG 2336
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480
 DB 2337 GCATTTGGTGGCTTGGCGAAGACGATCCCGCGCGCTTGGCGCATCAAACTTTGTATC 2396
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
 DB 2397 AGCATTTGCTTTGATACCGCGCTTCTGCGCTGATTTTACTAGCGACGAGTGGGTACG 2456
 QY 501 LeuAsnAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis 520
 DB 2457 CTCAATGACGACGCTGCTGCAAGACAGCAACATAGAGCGACGACGACGCTTGGCGCGCAC 2516

QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
 DB 2517 CTTTCGCGCTACACGAGCCCTGTATCGCGGACACGACGATCCGTCCGACCGACCGCGG 2576
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 DB 2577 CAAATCTATCAGGGCTTGGCGCATATGATTCGCTCGCGCAAGCAATCGCGCTTCGAC 2636
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
 DB 2637 GCGCGCAGCGCTGTGTACATTCACCAACCAACAGCACATCATCGGCTACATTCGCAAC 2696
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
 DB 2697 AATGGCTTTTGGCATTCGGTAACCTTCAGCGAATATCGCAAAACCGTTACCGCGCATAC 2756
 QY 601 LeuGlnAlaMetProPheIleAlaHisAspLeuIleGlyGlyValThrValSerLeuAsn 620
 DB 2757 CTGCAAGCATGCCCTTCAAGCGCACGACCTCATCGGTGGCAAACTGTTCAGCTTGAAT 2816
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpIleuGluIleAla 636
 DB 2817 CAGGATTTGACGCTTCAGCGCTTATCAGGTTCATGTGGCTCGAAATCGCC 2864
 RESULT 3
 AAA07380
 ID AAA07380 standard; DNA; 4173 BP.
 AC AAA07380;
 XX 30-JUN-2000 (first entry)
 DT
 XX
 DE Neisseria polysaccharea amylosucrase coding sequence.
 XX
 KW Amylosucrase; linear alpha-1,4 glucan production; fructose production;
 extracellular enzyme; cyclodextrin production; pure fructose syrup; ds.
 OS Neisseria polysaccharea.
 XX
 PN WO200014249-A1.
 PD 16-MAR-2000.
 XX
 PF 02-SEP-1998; 98WO-EP05573.
 XX
 PR 02-SEP-1998; 98WO-EP05573.
 XX
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
 XX
 PI Quanz M, Provart N;
 XX
 DR WPI: 2000-256695/22.
 DR P-PSDB; AAY81945.
 XX
 PT Novel amylosucrase polynucleotides and polypeptides obtained from
 Neisseria polysaccharea, used for production of alpha-1,4 glucans,
 fructose and cyclodextrins -
 XX
 PS Claim 1; Page 25-30; 36pp; English.
 XX
 CC This sequence encodes the Neisseria polysaccharea amylosucrase of the
 invention. The amylosucrase sequences are used for the in vitro
 production of linear alpha-1,4 glucans and fructose, as the enzyme is an
 extracellular enzyme. Amylosucrases may also be useful for the production
 of cyclodextrins. The amylosucrase sequences of the invention allow for
 the in vitro production of alpha-1,4 glucans and pure fructose syrup at
 low cost. Activated glucose derivatives or cofactors are not required.
 CC Immobilised cells can be at much higher densities than cells in liquid
 cultures, resulting in higher productivity.
 XX
 SQ Sequence 4173 BP; 926 A; 1196 C; 1118 G; 933 T; 0 other;
 Alignment Scores:

Pred. No.: 0 Length: 4173
 Score: 3408.00 Matches: 636
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-843-007a-2 (1-636) x AAA07380 (1-4173)

QY	1	MetLeuThrProThrGlnValGlyLeuLeuGlnTyrLeuLysThrArgileLeu	20
DB	1971	ATGTTGACCCCAACAGCAAGTCGGTTGATTTACAGTACTCAAAACAGCATCTTG	2030
QY	21	AspIleTyrThrProGlnArgIleGlyLeuLysSerGluAspTyrArgGlnPhe	40
DB	2031	GACATCTACACGCCCAACAGCGCCCGCATCGAAATCCGAGACGTGGCGCAGTTT	2090
QY	41	SerArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
DB	2091	TCGGCGCGCATGATACGCATTTCCTCCCAACCTGATGAACGACTCGACAGCGTGTACGGC	2150
QY	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTyrGlnSerTyr	80
DB	2151	AACACAGAGCCCTGCTGCTATGCTGAAATGCTGCTGGCGCAGCATGGCAAAAGCTAT	2210
QY	81	SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTyr	100
DB	2211	TCCCAACGCACATCATCTTAAAGATATCGATATCGCGCGGAAACCAACCCCGATTGG	2270
QY	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
DB	2271	ATTTTGTCCAAACAAAGTCGGCGCGTGTGCTAGCTTGATTTGTTTGGCGCGATTGG	2330
QY	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisIeu	140
DB	2331	AAGGCTTGAAGATAAAATTCCTATTTTCAAGAGCTTGTGTTGACTTATCTGCACCTG	2390
QY	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr	160
DB	2391	ATGCGCGTGTAAATGCTTGAAGCAAAAGCGAGCGCGCTATGCGGTACGACGCTAC	2450
QY	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu	180
DB	2451	CGCATGTTCATCCGCGCATCGGACCAATAGCGACTTGGCGAAGTCATTGCTGCGCTG	2510
QY	181	HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis	200
DB	2511	CACGACCGCGCATTTTCGCGCTGCTGATTTTATCTTCAACCACTTCACGAAACAC	2570
QY	201	GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220
DB	2571	GAATGGCGCAACGCTGCGCGCGCGGACCCGCTTTTCGACAAATTTCTACTATATTTTC	2630
QY	221	ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240
DB	2631	CCCGACCGCGGATGCGCGACCAATAGCGACCGCTGGCGGAAATCTTCCCGGACGAG	2690
QY	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrPheAsnSer	260
DB	2691	CACCGCGCGCTTCTCGCACTGGAAGACGCGCTGGGTGGACGACCTTCAATTC	2750
QY	261	PheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet	280
DB	2751	TTCCAAATGGGACCTTGAATACAGCAACCGCGGGTATTCGCGCAATGGCGGGCGAAATG	2810
QY	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr	300
DB	2811	CTGTTCTTGCCAACTTGGGCTTGACATCTGGGTATGATGATGGGTGCTTTATTTGG	2870
QY	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320
DB	2871	AAACAAATGGGGACCAAGCTGCAAAACCTGCGCGAGCGGACGCGCTCATCCGCGGCTTC	2930

RESULT 4

AAF61711
 ID AAF61711 standard; DNA; 6878 BP.
 XX AAF61711;
 AC AAF61711;
 XX
 DT 12-JUL-2001 (first entry)
 XX
 DE Expression vector pGEX-4T-1-AmsU containing amylosucrase DNA.

QY	321	AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340
DB	2931	AATCGCGTTATGCGTATTCGCGCGCGCGCGTCTTCAATCGGAAGCCATCGTCCAC	2990
QY	341	ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360
DB	2991	CCCGACCAAGTCGTCCTCAATACATCGGCGAGCAGCAATGCCAAATCGGTTCACAAACCCCTG	3050
QY	361	GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla	380
DB	3051	CAATGGCATGTTGTGGACACCCCTTGCACCGCGCGAAGTCAACCTGCTCCATCAGCGC	3110
QY	381	LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis	400
DB	3111	CTGACCTACCGCCCAACCTGCGCGACATACCGCTGGGTCAACTACGTCCGCGAGCCAC	3170
QY	401	AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr	420
DB	3171	GACGACATCGCGTGGAGCTTTCGCGATGAAGACCGGCAATCTGGGCGATAAGCGGTAC	3230
QY	421	AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg	440
DB	3231	GACCAACCGCAATTCCTCAACCGCTTCTCGTCAACCGTTTCGACGCGCAGCTTCGCTCGT	3290
QY	441	GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla	460
DB	3291	GGCGTACCGTTCCTCAACCGCTTTCGCGATGAAGACCGGCAATCTGGGCGATAAGCGGTAC	3350
QY	461	AlaLeuValGlyLeuAlaGlnAspProHisAlaValAspArgIleLysLeuLeuTyr	480
DB	3351	GCATTGCTCGCTTGGCGCAAGACGATCCCGACCGCTTGACCGCATCAACTCTTGCTAC	3410
QY	481	SerIleAlaLeuSerThrGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500
DB	3411	AGCATTTGCTTTGAGTACCGCGCTCTGCGCTGATTTACCTAGGCGCAGAGTGGGTACG	3470
QY	501	LeuAsnAspAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis	520
DB	3471	CTCAATGACGACGACACTTGGTTCGCAAGACAGCATTAAGAGCGACGACGCGCTTGGCGCGAC	3530
QY	521	ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540
DB	3531	CGTCCCGCTTACCAACGAGCCCTGTACGCGCAACGACGATCCGTCGACCGCGCGG	3590
QY	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
DB	3591	CAATCTATCAGGCGCTTGGCGCATATGATTTGCGTCCGCAAGCAATTCGCGCTTCGAC	3650
QY	561	GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn	580
DB	3651	GGCGGCGAGCTGTTTACATTTCAACCAACCAACGACATCATCGCTACATCCGCAAC	3710
QY	581	AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr	600
DB	3711	AATGCGCTTTTGGCATTCGTTAACTTCAGCGAATATCCGCAAAACCGTTTACCGCGCATAC	3770
QY	601	LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn	620
DB	3771	CTGCAAGCCATGCCCTTTCAGCGCGACGACCTCATCGGTGGCGAANAATGTGACGCTGAAT	3830
QY	621	GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla	636
DB	3831	CAGGATTTGACGCTTCAGCGCTTATCAGGTCTATGCTGGCTCGAAATCGCC 3878	

XX Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
 KW poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
 XX Unidentified.
 XX WO200125449-A2.
 XX PD 12-APR-2001.
 XX 04-OCT-2000; 2000WO-EP09695.
 XX 07-OCT-1999; 99DE-1048408.
 XX (AXIV-) AXIVA GMBH.
 XX Bengs H, Polakowski T, Held A, Gallert K;
 XX WPI; 2001-328330/34.
 XX Amylosucrase immobilized as fusion protein with anchoring sequence,
 XX useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
 XX

XX Example 2; Page 29-31; 38pp; German.
 XX This invention describes a novel amylosucrase (AS), immobilizable on a
 XX solid phase, which comprises a fusion protein (FP) of functional units
 XX of AS, an anchoring sequence, and optionally additional auxiliary
 XX sequences. The invention also describes (1) nucleic acid (1) encoding
 XX FP; (2) expression vector containing (1) and able to express FP in a
 XX host cell; Escherichia coli variants or fragments, of at least 8
 XX sequence, or its functional variants or fragments, of at least 8
 XX nucleotides that encodes an epitope, a high-affinity binding partner or
 XX GST (glutathione-S-transferase); (4) solid phase for immobilizing AS
 XX comprising glutathione-Sepharose; (5) combined, stable catalyst (A)
 XX comprising FP immobilized on Sepharose for production of
 XX poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).
 XX AS is used in production of poly(1,4-alpha-glucans) (II), useful for
 XX producing films, as food additives, as starting materials for
 XX cyclodextrins and as auxiliaries in pharmaceutical formulations.
 XX Immobilized AS makes possible efficient, inexpensive and continuous
 XX production of poly(1,4-alpha-glucans) (II), and it can be used
 XX repeatedly. Compared with known methods, specificity is improved
 XX (increased yield of (II) and reduced formation of palatinose) and
 XX reaction is complete within 24 hours, compared to 48-72 hours for batch
 XX methods. This sequence represents the expression vector construct
 XX pCEX-4T-1-AnSu which contains the amylosucrase (EC 2.4.1.4) described in
 XX the invention.
 XX

SQ Sequence 6878 BP; 1673 A; 1800 C; 1766 G; 1639 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 6878
 Score: 3408.00 Matches: 636
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0
 US-09-843-007A-2 (1-636) x AAF61711 (1-6878)

XX 1 MetLeuThrProGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20
 Db 945 ATGTTGACCCCAACGACGACGAGTGGTGGTTGATTTTACAGTACTCAAAACACGCAATCTG 1004
 XX 21 AspIleTyrThrProGlnGlnArgAlaGlyIleGluLysSerGluAspTyrArgGlnPhe 40
 Db 1005 GACATCTACACGCCCAACAGCGCGCGCATCGAAAATCGAAGACTGGCGGCAGTTT 1064
 XX 41 SerArgAcMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
 Db 1065 TCGCCCGCATGGATGACGATTTCCCAACTGATGAACGAACACTCGACAGCGGTGTACGGC 1124

QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuIleuAlaGlnAlaTpdGlnSerTyr 80
 Db 1125 AACAAAGAGCCCTGCTGCTATGCTGGAAATGCTGCTGGCGGCGGATGCGCAAGCTAT 1184
 QY 81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTyr 100
 Db 1185 TCCCAACGCAACTCATCTTAAAGATATCGATATCGCGCGGCAAAACAACCCGATGG 1244
 QY 101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
 Db 1245 ATTTTGTCCAAACAACAGTCCGCGCGGTGTGCTAGTGTGATTTGTTGCGCGGATTTG 1304
 QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuTyrTyrLeuHisLeu 140
 Db 1305 AAGGGCTTGAAGATAAATAATTCCTATTATTCAGAGCTTGGTTTGTATCTATCTGCACCTG 1364
 QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
 Db 1365 ATGCCCTGTGTTAAATGCCCTGAAGCAAAAGCGAGCGGGCTATCGGTTCAGCACTAC 1424
 QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu 180
 Db 1425 CGCGATGTCAATCGGCACTGGGCAATATAGCGACTTGGCGAGCTCATTTGTCGCTG 1484
 QY 181 HisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSerAsnGluHis 200
 Db 1485 CACGAAGCCGGCATTTCCGCCCTGCTGCGATTTTATCTTCAACACACACCTCCCAACGAAC 1544
 QY 201 GluTyrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
 Db 1545 GAATGGCGCAACGCTGGCGCGCGGACCGCGCTTTTCGACANTTCTACTATATTTTC 1604
 QY 221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
 Db 1605 CCGCACCGCGGATCCGACCAATACGACCGACCTCGCGGAAATCTTCCCGCACCA 1664
 QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrPheAsnSer 260
 Db 1665 CACCGCGCGGCTTCTCGCACTGGAAAGACGACCTGGGTGTGGACACCTTCATATTC 1724
 QY 261 PheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet 280
 Db 1725 TTCCAATGGGACTTGAATACAGCAACCGTGGGTATTCGCGCAATGGCGGCGCAATG 1784
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr 300
 Db 1785 CTGTTCTCTCCAACTTGGCGGTGACATCTCGGTATGGATGGGTGGCTTTATTTGG 1844
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
 Db 1845 AAACAAATGGGCAAAAGCTCGGAAACCTGCGCGAGCGGCGACGCCCTCATTCGCGCGTTC 1904
 QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
 Db 1905 AATGCCGTATGCGTATTTGCGCGCGCGCGCGCTGTCTTCAATCCGAAGCCATCGTCCAC 1964
 QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
 Db 1965 CCGGACCAAGTCGTCCATATCATCGGCGAGCAAGATGCCAAATCGGTACACCCCTG 2024
 QY 361 GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380
 Db 2025 CAATGGCATTTGTTGGAAACACCCCTTCCACGCGCGAAGTCAACCTCTCCATCAGCGC 2084
 QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis 400
 Db 2085 CTGACCTACCGCCACACCTGCCGAGCATACCGCTGGGTCACTACCTAGTCCGCGCAC 2144
 QY 401 AspAspIleGlyTyrThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr 420
 Db 2145 GACGACATCGCTGGACGTTTGGCGATGAAGACGCGGCATATCTGGGCATGAAGCGGCTAC 2204

QY 421 AspHisArgGlnPheLeuAsnArgPheValAsnArgPheAspGlySerPheAlaArg 440
 Db 2205 GACCACCGCAATTCCTCAACCGCTCTTCGTCAACCGTTTCGACGGAGCTTCGCTCGT 2264
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
 Db 2265 GCGGTACCGCTTCCATACCAACCAACACAGCGCACTGCGGTGTCAGTGGTACAGCGCG 2324
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAsnArgTleLysLeuLeuTyr 480
 Db 2325 GCATGTGCTGGTGGCGCAAGCGATCCCGCGGTGACCGCATCAAACTCTTGAC 2384
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
 Db 2385 AGCAATTCCTTTCAGTACCGCGCTGCGCTGATTTACCTAGGCGACGAAGTGGTACG 2444
 QY 501 LeuAsnAspAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis 520
 Db 2445 CTCATATGACGACACTGGTGGCAAGACCAATATAGAGCGACGACAGCGGTGGCGGAC 2504
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
 Db 2505 CGTCGCGCTCAACGAAGCCCTGTACGCGCAACGCAACGATCCGTTCGACCGCGG 2564
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 Db 2565 CAAATCTATCAGGGCTTGGCGCATATGATTCGCGTCCGCCCAAGCAATCCGCGCTTCGAC 2624
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
 Db 2625 GCGCGCAGGCTGTTACATTCACACCAACACACACACATCATCGGTACATCCGCAAC 2684
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
 Db 2685 AATGCGCTTTTGGCAATTCGTTAACTTCAGCGAATATCCGCAACGTTTACCGCGCATACC 2744
 QY 601 LeuGlnAlaValProPheLysAlaHisAspLeuIleGlyLysThrValSerLeuAsn 620
 Db 2745 CTGCAAGCCATGCCCTTCAGCGCGCAGCACTTCATCGGTGGCAAACTGTACGCTGAAT 2804
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla 636
 Db 2805 CAGGATTCGACCTTCAGCCCTATCAGGTTCATGTGGCTCGAATCGCC 2852
 RESULT 5
 AAF61710
 ID AAF61710 standard; DNA; 1910 BP.
 AC AAF61710;
 DT 12-JUL-2001 (first entry)
 DE Amylosucrase PCR derived DNA fragment AmSu5.
 XX
 KW Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
 XX poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
 OS Unidentified.
 XX
 FN W0200125449-A2.
 XX
 PD 12-APR-2001.
 XX
 PF 04-OCT-2000; 2000WO-EP09695.
 XX
 PR 07-OCT-1999; 99DE-1048408.
 XX
 PA (AXIV-) AXIVA GMBH.
 XX
 PI Bengs H, Polakowski T, Held A, Gallert K;
 XX WPI; 2001-328330/34.
 XX

PT Amylosucrase immobilized as fusion protein with anchoring sequence,
 PT useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
 XX
 PS Claim 3; Page 28; 38pp; German.
 XX
 CC This invention describes a novel amylosucrase (AS), immobilizable on a
 CC solid phase, which comprises a fusion protein (FP) of functional units
 CC of AS, an anchoring sequence, and optionally additional auxiliary
 CC sequences. The invention also describes (1) nucleic acid (I) encoding
 CC host cell; Escherichia coli containing the vector of (2); (3) anchoring
 CC sequence, or its functional variants or fragments, of at least 8
 CC nucleotides that encodes an epitope, a high-affinity binding partner or
 CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS
 CC comprising glutathione-Sepharose; (5) combined, stable catalyst (A)
 CC comprising FP immobilized on Sepharose for production of
 CC poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).
 CC AS is used in production of poly(1,4-alpha-glucans) (II), useful for
 CC producing films, as food additives, as starting materials for
 CC cyclodextrins and as auxiliaries in pharmaceutical formulations.
 CC Immobilized AS makes possible efficient, inexpensive and continuous
 CC production of poly(1,4-alpha-glucans) (II), and it can be used
 CC repeatedly. Compared with known methods, specificity is improved
 CC (increased yield of (II) and reduced formation of palatinose) and
 CC reaction is complete within 24 hours, compared to 48-72 hours for batch
 CC methods. This sequence encodes a PCR derived amylosucrase (EC 2.4.1.4)
 CC AmSu5 fragment described in the invention.
 SQ Sequence 1910 BP; 447 A; 595 C; 474 G; 387 T; 7 other;

Alignment Scores:

Pred. No.: 0 Length: 1910
 Score: 3360.00 Matches: 626
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.84% Mismatches: 0
 Query Match: 98.59% Indels: 0
 DB: 22 Gaps: 0

US-09-843-007A-2 (1-636) x AAF61710 (1-1910)

QY 10 LeuIleLeuGlnTyrLeuLysThrArgIleLeuAspIleTyrThrProGluGlnArgAla 29
 Db 10 ATGATTTTACAGTACTCTCAAAACACAGCATCTTGGACATCTACACGCCGACAGCGGCC 69
 QY 30 GlyIleGluLysSerGluAspTyrArgGlnPheSerArgArgMetAspThrHisPhePro 49
 Db 70 GGCATCGAAAATCCGAAGACTCGCGCGCAGTTTTCGCGCGCATGCGATATTCCTCC 129
 QY 50 LysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsnGluAlaLeuLeuProMetLeu 69
 Db 130 AACTGATGACGNACTCGACGGTGTACGGCAACACGACGACCTGCTGCTATGCTG 189
 QY 70 GluMetLeuLeuAlaGlnAlaTyrGlnSerTyrSerGlnArgAsnSerSerLeuLysAsp 89
 Db 190 GAAATGCTGTGCGCGCAGGATGCAAAAGCTATTCCACAGCGCAACTCATCTCTAAAGAT 249
 QY 90 IleAspIleAlaArgGluAsnAsnProAspTyrIleLeuSerAsnLysGlnValGlyGly 109
 Db 250 ATCGATATCGCGCGCGAATAACACCCCGATTTGATTTTGTCCAAACAAACAGTCGCGCG 309
 QY 110 ValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyr 129
 Db 310 GTGTGCTACCTGATTTGTTTTCGCGCGATTTGAAGGGCTTGAAGAGATAAAATTCCTTAT 369
 QY 130 PheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGluGly 149
 Db 370 TTTCAGAGCTTGGTATTGACTTATCTGACCTGATCGCGCTGTGTTAAATGCCCTGAAGGC 429
 QY 150 LysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThr 169
 Db 430 AAAAGCGACGCGGCTATGCGGTACGAGCTACCGCGATCTCAATCCGCACTGGGCACA 489

QY 170 IleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIleSerAlaValVal 189
 DB 490 ATAGGCGACTTGGCGAAGTCATGTGCGCTGCACGAAGCGCGCATTTCCGCGCTGTC 549
 QY 190 AspPheIlePheAsnHisThrSerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGly 209
 DB 550 GATTTTATCTTCAACCACTCCACGACCTCCACGACAGATGGCGCAAGCTGCGCGCGCGC 609
 QY 210 AspProLeuPheAspAsnPheTyrTrpIlePheProAspArgArgMetProAspGlnTyr 229
 DB 610 GACCCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATAC 669
 QY 230 AspArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGlu 249
 DB 670 GACCGCACCTTGGCGCAATCTTCCCGGACGACACCGCGCGCTTCTCGGACCTGGAA 729
 QY 250 AspGlyArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269
 DB 730 GACGGACGCTGGGTGTGGACGACCTTCAATTCCTTCCAAATGGGACTTGAATTACAGCAAC 789
 QY 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
 DB 790 CCGTGGGTATTCGCGCAATGGCGGCGGAAATGCTGTTCTTCCCAACTTGGCGGTGAC 849
 QY 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309
 DB 850 ATCTCGGTATGATCGGTGCTGCTTTATTTGGAAACAATGGGACAAAGCTGGCAAAAC 909
 QY 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaPro 329
 DB 910 CTGCGCGACGCGCACCCCTCATCCCGCGGTTCATATGCGGTATGCGGTATTCGCGCGCC 969
 QY 330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349
 DB 970 GCGGTGTTCTTCAATTCGAGCAATCGTCCACCCGACCAAGTCGTCCAATACATCGG 1029
 QY 350 GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeu 369
 DB 1030 CAGGACGAATGCGAAATCGGTTACAAACCCCTCGCAATGGCATTTGTGGAAACACCTT 1089
 QY 370 AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389
 DB 1090 GCCACGCGGAGTCACTGCTCTCATCAGCGCTGACTACCGCACCAACTGCGCGGAG 1149
 QY 390 HisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPheAlaAsp 409
 DB 1150 CATACGCGCTGGGTCAACTACGTCGCGCAGCACGACATCGGCTGGACGTTTGCCTGAT 1209
 QY 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 429
 DB 1210 GAAGACGCGGCATATCTGGGCATAGCGGTAGCACCGCGCAATTCCTCAACGCTTC 1269
 QY 430 PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer 449
 DB 1270 TTGCTCAACGTTTCGACGCGACCTGCTCGTGGGTACCTGCTCCATACCAACCAAGC 1329
 QY 450 ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp 469
 DB 1330 ACAGGCGACTCGCGTGCAGTGATACAGCGCGGCAATGGTGGCTTGGCGCAAGACGAT 1389
 QY 470 ProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu 489
 DB 1390 CCCCACGCGGTGACCGCATCAACTCTTGTACAGCATTTGCTTGTAGTACCGGCGTCTG 1449
 QY 490 ProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspTrpSerGlnAsp 509
 DB 1450 CCGCTGATTTACCTAGCGCACGAAGTGGGTGATCGCTCAATGACGACGACCTGGTGGCAAGAC 1509
 QY 510 SerAsnLysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGlnAlaLeuTyr 529
 DB 1510 AGCAATATAGAGGACGACAGCGGTGGCGGACCGTCCGCGCTACCAAGAGCCCTGTGAC 1569
 QY 530 AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMet 549

DB 1570 GCGCAACCGCAACGATCCGTCGACCGCGCAAAATCTATCAGGGCTTCGCCCATATG 1629
 QY 550 IleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThr 569
 DB 1630 ATTGCGCTCCGCCCAAGCAATCCGCGCTTCGACGCGGCGGCTGGTTACATTCACACC 1689
 QY 570 AsnAsnLysHisIleIleGlyTyrIleArgAsnAsnAlaLeuLeuAlaPheGlyAsnPhe 589
 DB 1690 AACACAAGCACATCATCGGTACATCCGCAACCAATGCGCTTTTGGCAATTCGGTAACCTC 1749
 QY 590 SerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis 609
 DB 1750 AGCGAATATCGCAACCGTTACCGCGATACCTTCGACGACATGCCCTTCAGGCGGAC 1809
 QY 610 AspLeuIleGlyGlyLysThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
 DB 1810 GACCTCATCGGTGGCAAACTGTGACGCTGAATCAGGATTTGACGCTTCAGCCCTATCAG 1869
 QY 630 ValMetTrpLeuGluIleAla 636
 DB 1870 GTCATGTGCTCGAAATCGCC 1890
 RESULT 6
 ID AAF61712 standard; DNA; 6851 BP.
 AC AAF61712;
 DT 12-JUL-2001 (first entry)
 DE Expression vector pGEX-4T-1-AmsU5 containing amylosucrase DNA.
 KW Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
 KN poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
 XX Unidentified.
 FT Key Location/Qualifiers
 CDS 258..2828
 /*tag= a
 /product= "AmsU5"
 FN WO200125449-A2.
 PD 12-APR-2001.
 PF 04-OCT-2000; 2000WO-EP09695.
 PR 07-OCT-1999; 99DS-1048408.
 PA (AXIV-) AXIVA GMBH.
 PI Bengs H, Polakowski T, Held A, Gallert K;
 DR WPI: 2001-328330/34.
 DR P-PSDB; AA570883.
 PT Amylosucrase immobilized as fusion protein with anchoring sequence,
 PT useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
 PS Claim 6; Page 31-34; 38pp; German.
 XX This invention describes a novel amylosucrase (AS), immobilizable on a
 CC solid phase, which comprises a fusion protein (FP) of functional units
 CC of AS, an anchoring sequence, and optionally additional auxiliary
 CC sequences. The invention also describes (1) nucleic acid (I) encoding
 CC FP; (2) expression vector containing (I) and able to express FP in a
 CC host cell; Escherichia coli containing the vector of (2); (3) anchoring
 CC sequence, or its functional variants or fragments, of at least 8
 CC nucleotides that encodes an epitope, a high-affinity binding partner or
 CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS

Qy	610	AspLeuIleGlyGlyThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGln	629
Db	2745	GACCTCATCGTGGCAAAACGTGACGCTGAATCAGGATTTCAGCGCTTCAGCCCTATCAG	2804
Qy	630	ValMetTyrLeuGluIleAla	636
Db	2805	GTCATGTGGCTCGAATCGCC	2825
RESULT 7			
ID	AA09860		
XX	AA09860	standard; DNA; 2883 BP.	
AC	AA09860;		
XX	08-APR-1996	(first entry)	
DE	Neisseria polysaccharea	amylosucrase DNA sequence.	
XX	Amylosucrase; enzyme; amylose, linear 1,4-glucan; transgenic plant;		
KW	EC-2.4.1.4; crop improvement; ss.		
OS	Neisseria polysaccharea.		
FT	Key	Location/Qualifiers	
FT	5'UTR	1..938	
FT	sig_peptide	/*tag= a	
FT	CDS	939..986	
FT		/*tag= b	
FT		939..2780	
FT		/*tag= c	
PN	W09531553-Al.		
XX	23-NOV-1995.		
XX	18-MAY-1995;	95MO-EP01893.	
PR	22-DEC-1994;	94DE-4447388.	
PR	18-MAY-1994;	94DE-4417879.	
XX	(GENB-) INST GENBIOLOGISCHE FORSCHUNG.		
PA	Buettcher V, Kossmann J, Welsh T;		
PI	WPI; 1996-010938/01.		
DR	P-PSDB; AAR88386.		
XX	DNA encoding amylo:sucrase from Neisseria polysaccharea - for prodn.		
FT	of linear 1,4-glucan(s), esp. amylose, from sucrose.		
XX	Claim 1; Page 39-42; 56pp; English.		
CC	This DNA sequence encodes an amylosucrase which allows the		
CC	synthesis of linear alpha-1,4-glucans from the substrate sucrose		
CC	by bacteria, fungi and plants, or in cell-free systems. This		
CC	sequence may be expressed recombinantly.		
XX	Sequence 2883 BP; 661 A; 852 C; 758 G; 612 T; 0 other;		
Alignment Scores:			
Pred. No.:	0	Length:	2883
Score:	3141.50	Matches:	621
Percent Similarity:	98.11%	Conservative:	3
Best Local Similarity:	97.64%	Mismatches:	11
Query Match:	92.18%	Indels:	9
DB:	17	Gaps:	1
US-09-843-007A-2 (1-636) x AA09860 (1-2883)			
Qy	1	MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu	20
Db	939	ATGTTGACCCCGGAGGTCGGTTTGATTTCAGTACCTCAAAACACGACATCTTG	998

Qy	21	AspIleTyrThrProGlnArgAlaGlyIleGluLysSerGluAspTyrArgGlnPhe	40
Db	999	GACATCTACAGCGCCGAAACAGCGCGCGCATCGAAATAATCCGAAGACTCGCGCAGTTT	1058
Qy	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
Db	1059	TCGCGCCGATGATACGCAATTTCCCAACTGATGAACGAACACTCGACAGCGTGTACGGC	1118
Qy	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTyrGlnSerTyr	80
Db	1119	AACAACGAGCCCTGCTGCTATGCTGGAATGCTGCTGGCGCAGGATGCGAAAGCTAT	1178
Qy	81	SerGlnArgAsnSerSerLeuLysAspIleAlaArgGluAsnAsnProAspTyr	100
Db	1179	TCCCAACGCACTCATCTTAAGATATCGATATCGCGCGGAGAAACAACCCGATTGG	1238
Qy	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
Db	1239	ATTGTTGCCAACAACAAAGTCGGCGCTGTGCTACGTTGATTGTTTGGCGCGGATTG	1298
Qy	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu	140
Db	1299	AAGGCTTGAAAGATAAATTCCTTATTTTCAGAGCTTGCTTGACTTATCTGCACCTG	1358
Qy	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr	160
Db	1359	ATGCGCTGTTTAAATGCCCTGAAGGCAAGAGCGCGCTATGCGTTCAGCACTAC	1418
Qy	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu	180
Db	1419	CGGATGTCAATCCGCACTGGGCACAATAGCGACTTGGCGAGTCAATTGCTGGCTG	1478
Qy	181	HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis	200
Db	1479	CACGAATCGC - ATTTCGCGCTGCTCGATTATTTATCTTCAACCACTCCCAAGCAAC	1536
Qy	201	GluTyrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220
Db	1537	GAATGGCG - CAACGCTGC - --GCCGGCGACCCGCTTTTCGCAATTTCTACTATATTTC	1592
Qy	221	ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240
Db	1593	CCCGACCGCGGATGCCGACCAATACGACCGACCTTGGCGGAAATCTTCCCGACGAG	1652
Qy	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrPheAsnSer	260
Db	1653	CACCGCGCGCTTCTCGCAACTCGAAGACGCGCTGGGTGTGGACGACCTTCAATTCC	1712
Qy	261	PheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet	280
Db	1713	TTCANATGGGACTTGAAATACAGCAACCCGTGGGTATTTCG - GCAATCGCGGCGAAATG	1771
Qy	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr	300
Db	1772	CTGTTCCTTGGCAACTTGGGGGTTGACATCTCGTATGATGCGGTTTTCCTTTATTGG	1831
Qy	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320
Db	1832	AAACAAATGGGACAAAGCTGCAAAACCTGC - GCAGGCGACGCTCATTCGCGGCTTC	1889
Qy	321	AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340
Db	1890	AATGCGGTATGCGTATTTCGCGCGCCGCTGTTCTTCAAAATCGGAAGCCATCTGCCAC	1949
Qy	341	ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360
Db	1950	CCCGACCAAGTCGTCCAATATCGGGCAGGACGAATGCAAAATCGTTACCAACCCCTTG	2009
Qy	361	GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla	380
Db	2010	CAATGGCATTTGTTGTGGACACCCCTTCCACGCGCGAAGTCAACCTGCTCCATCAGCG	2069

QY 381 LeuThrTyArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400
 Db 2070 CTGACCTACCGCCACAACTGCCGAGCATATCCGCTGGGTCACTAGCTCCGAGCCAC 2129
 QY 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr 420
 Db 2130 GAAGCATCGCTGGACGTTTGGCCATGAACGCGGCGATATCTGGGCGATAGCGGGCTAC 2189
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
 Db 2190 GACCACCGCCATCTCTCAACCGCTTCTTCGTCACCGTTTCGACGGACGCTCGCTCGT 2249
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
 Db 2250 GCGGTACCGTTTCCATACACCCAGACACGCGGACTCGCGTGTCTGAGTGTACGCGCG 2309
 QY 461 AlaLeuValGlyLeuAlaGluAspProHisAlaValAspArgIleLysLeuLeuTyr 480
 Db 2310 GCATTTGGTGGCTTGGCGGAGACGATCCCGCGCGTTGACCGATCAAACTCTTTGTAC 2369
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
 Db 2370 AGCATTGCTTGGTACCGCGGCTCTCGCGCTGATTACCTAGCGGACGAAAGTGGGTACG 2429
 QY 501 LeuAsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis 520
 Db 2430 CTCATATGACGACGACTGGTGCCAGC-AGCAATAGACGCGACGACGCGGTGGGC-CAC 2487
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
 Db 2488 CGTCGCGCTACACGAGCCCTGTACGGGCAACCGACGATCCGTGACCGCGCGCGC 2547
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 Db 2548 AA-ATCTATCATCGGCTTGGCCATATGATTCGCGTCCGCCAAAGCAATCCGCGTTCAC 2606
 QY 561 GlyGlyArgGluValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
 Db 2607 GCGCGAGCTGGTTACATTTACACCAACCAACAGCAATCATCGGTACAT-CGCAAC 2665
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGlnTyrProGlnThrValThrAlaHisThr 600
 Db 2666 AATGCGCTTTTGGCATTCGGTAATTCAGCGAATATCCGCAACCGTTTACCGCGCATACC 2725
 QY 601 LeuGlnAlaMetProPheLysAlaHisAspIleuIleGlyGlyLysThrValSerLeuAsn 620
 Db 2726 CTCGAAGCGCATGCCCTTCAAGCGCGACGACCTCATCGGTGGCAAACTGTGAGCGCTGAAT 2785
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
 Db 2786 CAGGATTTACGCTTCAGCCCTATCAGGTCTATGCTGCTCGAATCGCC 2833
 RESULT 8
 AAT11179
 ID AAT11179 standard; DNA; 2883 BP.
 AC AAT11179;
 XX
 XX
 DT 20-JUN-1996 (first entry)
 DE
 DE Neisseria polysaccharea amylosucrase gene.
 KW Amylosucrase; bacteria; fungi; plants; detection; transformation;
 KW linear; alpha-1,4-glucans; amylose; sucrose; colourless;
 KW odourless; tasteless; non-toxic; biodegradable; self-sustaining;
 KW films; fibres; textiles; paper-making; glass-fibre; tablet binder;
 KW food thickener; sound proofing; flow properties; paraffin oils;
 KW organic compound inclusion; chromatographic separation;
 KW cyclodextrins; ss.
 XX
 OS Neisseria polysaccharea.
 XX
 XX Key Location/Qualifiers
 FH

FT CDS 939..2783
 FT /*tag= a
 XX DE4417879-A1.
 XX
 XX 23-NOV-1995.
 XX
 XX 18-MAY-1994; 94DE-4417879.
 XX
 XX 18-MAY-1994; 94DE-4417879.
 XX
 XX (GENB-) INST GENBIOLOGISCHE FORSCHUNG.
 XX
 XX Buettcher V, Kossmann J, Welsh T;
 XX P-PSDB; AAR88633.
 DR WPI: 1996-000447/01.
 DR
 DR New DNA sequence encoding amylo:sucrase of Neisseria - and
 FT transformed plant, bacteria and fungi able to produce linear
 FT alpha-1,4-glucon(s), esp. amylose, in practically pure form
 FT
 PS Claim 6; Pages 27-33; 42pp; German.
 XX
 CC The N. polysaccharea DNA sequence AAT11179, which encodes AAR88633
 CC amylosucrase (ASA), can be used to produce bacteria, fungi and
 CC plants that express ASA, and to detect and isolate related DNA
 CC from other organisms. Transformed plants which express ASA are
 CC able to produce linear alpha-1,4-glucans, specifically amylose
 CC from sucrose, which can be used to produce colourless,
 CC odourless, tasteless, nontoxic, biodegradable, self-sustaining
 CC films or fibres, e.g. for use in the food, textile, paper-
 CC making and glass-fibre industries. Amylose can also be used as
 CC a binder for tablets, thickener for food, in sound proofing
 CC panels, to improve flow properties in paraffin-based oils, for
 CC inclusion of organic cpds., in chromatographic sepn. and as a
 CC starting material for cyclodextrins.
 XX
 SQ Sequence 2883 BP; 561 A; 852 C; 758 G; 612 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 2883
 Score: 3141.50 Matches: 621
 Percent Similarity: 98.11% Conservative: 3
 Best Local Similarity: 97.64% Mismatches: 11
 Query Match: 92.18% Indels: 9
 DB: 17 Gaps: 1
 US-09-843-007A-2 (1-636) x AAT11179 (1-2883)
 QY 1 MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20
 Db 939 ATGTTGACCCCGACGAGTGGTGGTTGATTTACGTACTCAAAACACGATCTTG 998
 QY 21 AspIleTyrThrProGlnArgAlaGlyIleGlyLysSerGlnAspTrpArgGlnPhe 40
 Db 999 GACATCTACAGCCCGAAGCAGCGCGCGCATCGAAAAATCCGAAGACTGGCGGAGTTT 1058
 QY 41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAsnSerValTyrGly 60
 Db 1059 TCGCGCGCATGGATACGCAATTTCCCAACTGATGAACGAACTCGACGGGTGTACGCG 1118
 QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr 80
 Db 1119 AACAAACGAGCCCTGCTGCTATCTCGAATGCTGCTGGCGGAGGATGCAAGCTAT 1178
 QY 81 SerGlnArgAsnSerSerLeuLysAspIleAlaArgGluAsnAsnProAspTrp 100
 Db 1179 TCCCAACGCACTCATCTTAAAAAGATATCGATATCGCGCGGAAACCAACCCCGATTGG 1238
 QY 101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
 Db 1239 ATTTTGTCCAAACAAACAGTGGCGGGGTGCTGCTACGTTGATTGTTGCCGCGGATTG 1298

QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
Db 1299 AAGGCTTTGAAAGATAAAATTCCTATTATTTCAAGAGCTTGGTTGACTTATCTGCACCTG 1358
QY 141 MetProLeuPheLysCysProGluGlySerAspGlyGlyTyrAlaValSerSerTyr 160
Db 1359 ATGCCGCTGTTTAAATGCTGTAGAGCAAAAGCGACGGCGCTATGCGGTGACGACGTAC 1418
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuA:rgGluValIleAlaAlaLeu 180
Db 1419 CGCGATGTCATTCGGCACTGGGCAATAGCGACTTGGCGAGTCAATGCTGCGCTG 1478
QY 181 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200
Db 1479 CACGAATGCG--ATTTCGCGCGTGTGCGATTATCTTCAACCACTCTCCAGCAACAC 1536
QY 201 GluTyrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
Db 1537 GAATGGCG-CAACGGCTGC--GCCGGCGACCGCTTTTCGACAAATTCTACTATATTTC 1592
QY 221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
Db 1593 CCCGACCGCGGATGCCGACCAATACGACCCCTGCGGGAATCTTCCCGGACACAG 1652
QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrPheAsnSer 260
Db 1653 CACCGCGGCGCTTCTCGCACTGAGAGACGACGCTGCGTGTGACGACCTTCAATTCC 1712
QY 261 PheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyLysMet 280
Db 1713 TTCCATGGGACTTGAATTACAGCAACCGTGGTATTTCGC-GCAATGGCGGCGGCAATG 1771
QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr 300
Db 1772 CAGTCTTTCGCAACTTGGCGGCTGTGCATCTTGGTATGATGATGCGTTGCCTTTATTTGG 1831
QY 301 LysGlnMetGlyThrSerCysGluAsnLeuP:roGlnAlaHisAlaLeuIleArgAlaPhe 320
Db 1832 AAACAATGGGACAGCTGCAAAACCTGCG--GCAGCGACGCGCTCATCCGCGCGTTC 1889
QY 321 AsnAlaValMetArgIleAlaProAlaValPhePheLysSerGluAlaIleValHis 340
Db 1890 AATGCCGTTATGGGTATTCGCGGCGCGGCTGTCTTCAATTCGAGCCATGCTCCAC 1949
QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
Db 1950 CCGGACCAAGTCGTCCAATACATGCGGACGAGCAATGCCAAATCGTTTACAACCCCTG 2009
QY 361 GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAla 380
Db 2010 CAAATGGCAATTGTTGGCAACCTTGCACGCGCGAAGTCAACCTGCTCCATCAGCGG 2069
QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis 400
Db 2070 CTGACCTACCGCCACACCTGCGGAGCATACCGCTGGGTCAACTAGCTGCGGACCCAC 2129
QY 401 AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuClyIleSerGlyTyr 420
Db 2130 GACGACATCGCTGGACGTTTGCATGAGACGCGGCATATCTGGGCATAGCGGCTAC 2189
QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
Db 2190 GACACCGCAATTCCTCAACCGCTTCTGTCACACCGTTTCGCGGACGTTGCTGCT 2249
QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
Db 2250 GCGGTACCGTTCGAATACACACCCAGCAGCGGAGCTGCGGTGTCAGTGTACAGCGCGG 2309
QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480
Db 2310 GCATTGTGCGCTGGCGCAAGACGATCCCCACGCGGTTGACCGCATCAAACTCTGTAC 2369

QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
Db 2370 AGCATTCCTTTAGTACCGCGCTCTGCGCTGATTACCTAGCGGACGAAGTGGGTACG 2429
QY 501 LeuAsnAspAspAspTyrPheGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis 520
Db 2430 CTCATGACGACGACTGGTCCCAAGC-AGCAATAAGAGCGACGACGCCCTTGGGC-CAC 2487
QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
Db 2488 CGTCCGCGCTACACGAAGCCCTGTACGCGCAACCGAACGATCCGTGACCGGACCGCGC 2547
QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
Db 2548 AA-ATCTATCAGGCGCTTGGCCATATGATTGCGCTCGCGCAAAACAAATCCGCGCTTCAC 2606
QY 561 GlyClyArgLeuValThrPheAsnThrAsnLysHisIleIleGlyTyrIleArgAsn 580
Db 2607 GCGGCGAGGCTGGTTACATTCACACCAACCAACGACATCATCGGCTACAT-CGCAAC 2665
QY 581 AsnAlaLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
Db 2666 AATCGCTTTTGGCATTCGGTAACCTCAGCGAATATCCGAAACCGTTACCGCGCATACC 2725
QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
Db 2726 CTGCAAGCCATGCGCTTCAAGGCGACGACCTCATCGGTGGCAAACTGTGCGCTGAAT 2785
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla 636
Db 2786 CAGGATTTGAGCTTTCAGCGCTATCAGCTCATGTGGTCGAAATCGCC 2833
RESULT 9
AAS59546/c
ID AAS59546 standard; DNA; 29559 BP.
AC AAS59546;
XX 13-FEB-2002 (first entry)
DT
DE
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.
OS Propionibacterium acnes.
PN WO200181581-A2.
PD 01-NOV-2001.
PF 20-APR-2001; 2001WO-US12865.
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX (CORI-) CORIXA CORP.
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX Claim 1; SEQ ID No 41; 1069pp; English.
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding

CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
 CC associated DNA sequences are used in the treatment, prevention and
 CC diagnosis of medical conditions caused by P. acnes. The disorders include
 CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
 CC in infections of bone, joints and the central nervous system, however it
 CC is particularly involved in the inflammatory lesions associated with acne
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a
 CC patient comprises contacting a sample with a binding agent that binds to
 CC the proteins of the invention and determining the amount of bound protein
 CC in the sample. The polypeptides may be used as antigens in the production
 CC of antibodies specific for P. acnes proteins. These antibodies can be
 CC used to downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
 CC polypeptides shown in AAU49884-AAU50191 and AAU67524-AAU67527.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 29559 BP; 5419 A; 8656 C; 9368 G; 6114 T; 2 other;

Alignment Scores:

Pred. No.:	3.17E-110	Length:	29559
Score:	1164.00	Matches:	245
Percent Similarity:	57.56%	Conservative:	113
Best Local Similarity:	29.29%	Mismatches:	230
Query Match:	34.15%	Indels:	34
DB:	23	Gaps:	12

US-09-843-007A-2 (1-636) x AAS59546 (1-29559)

QY	38	ArgGlnPheSerArgAcgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSer	57
DB	22070	CAGTCTTCGACCTCCGCTGGGAGCACTACTACCTGATCTATGGATCGGCTGAGCAGG	22011
QY	58	ValTyrGlyAsn-----AsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGln	75
DB	22010	GTCTAGCGAGACCGCGCGGATGAACCTCGCAGAGATCCGCGCATCTCTCTAAACGG	21951
QY	76	AlaTrpGlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGlu	95
DB	21950	TGT-----GCCAGACGCTCTGATGATCTCAAGCGCTCGACGAGCCCGGCTC	21903
QY	96	AsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeu	115
DB	21902	CTGGAGCCCGACTGGCTTCAACACCGACGATGATCGGTATGCACTATACGACCAT	21843
QY	116	PheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeu	135
DB	21842	TTCCTCGGCACCTCAAGGATATAAGCGATCACTCGATCACCTCTGCGACATGGGTG	21783
QY	136	ThrTyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyr	155
DB	21782	CGTATCTGCATCTCATCCCTCTGCAACCTCGCCANGTACTGACGATGGTGGCTAT	21723
QY	156	AlaValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGlu	175
DB	21722	CGCGTGGGACCATCGCACTATCCGACCGATCTGGGTACTACTGATGACCTAGCTGAC	21663
QY	176	ValIleAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHis	195
DB	21662	CTCAGCGGTACCTCGCGCCACCGCATATCTCTGTCATGACCTCATGTTAACCCAC	21603
QY	196	ThrSerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsn	215
DB	21602	GTAGCGCGGACGATGATGGCTCGAGAGCTCGGGCTGTCTCAACAGAGTACCGCGAC	21543
QY	216	PheTyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu	235
DB	21542	TATTTCCATATTCTGTGACTCAAGATGAGGTTCACGCTCGGAGAAACCTTCGCGAC	21483

QY	236	IlePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg---TrpVal	254
DB	21482	GTITTTTCGGACCTTCGGCGCATGGCAATTTCCCTGGGACACGACCTGCCAGGTGGGTA	21423
QY	255	TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArg	274
DB	21422	TGGCGGACCTTCAACGAAATTCAGTGGATCTCACTGGGTACCTGACGTTTTCGC	21363
QY	275	AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp	294
DB	21362	GAAATTCCTCGACCTGATCGCTGTCTGGCCATCGCGGGTCGAGGTGTTCGCGCTGAC	21303
QY	295	AlaValAlaPheIleTyrPheLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis	314
DB	21302	GCCATCGCTTCATCTGAGAAACCTCGGCACCAACTGTGAGAACCTTCGGAATTCAC	21243
QY	315	AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys	334
DB	21242	GACATCACTCAGTCATTGCTGAGCAATACGATCGTCGCGCGCGCTTCATG	21183
QY	335	SerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGlnAspGlu-----	352
DB	21182	GCGATGCCATCGTGGCCCGGACGATCTCACGGATATTTGCGACCGCGGCGCATGG	21123
QY	353	-----CysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeu	369
DB	21122	GGAAAGTCTCGCAGCATGATCTATCAACACAGCTCATGGTGCAGCTGTGGAGCGCCTA	21063
QY	370	AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu	389
DB	21062	GCTACCGCGACGTCAGCTCATGGAACACAGCGTTCAGTCGGACCGCGCAACCCCTCG	21003
QY	390	HisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPheAlaAsp	409
DB	21002	ACACAACTTGGCGCCACTAGCTGATGTCACGACGACATCGGATGGACCGTGCATGAT	20943
QY	410	GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe	429
DB	20942	GCGACGCGCGTAAACCTGCGCTTGAACCGCTGCGCCACCGCGAGTTCCTCTCGACTTC	20883
QY	430	PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer	449
DB	20882	TATTCGGGACATTCCTCCCGATCTCTGCGCGGTCTGTCTTCCGACGACACCCCGGT	20823
QY	450	ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGly-----Leu	465
DB	20822	ACCGGACCGTTCGAATTAGTGGCTCTTAGCAAGCGCTGGCGGCTGGAATCTGCGCTG	20763
QY	466	AlaGlnAspAspProHisAlaValAsp-----ArgIleLysLeuLeuTyrSer	481
DB	20762	GAGTCCGACACCGACGCTGCTGTGACGCTGCCATCGCCGCGATCGTATGCTGCACACC	20703
QY	482	IleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeu	501
DB	20702	GCGNTCTCGGTACGGCGGAGTAGTACCACTAATCTGGATGGAGACGAGGTTCGGATGCTC	20643
QY	502	AsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHisArg	521
DB	20642	AAC---GACGACTGGCAACGTCATCGCGGTCTGCGGACGACACCGCTGGTCCATCGA	20586
QY	522	ProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsn---AspProSerThrAlaAlaGly	540
DB	20585	CCTATGATGAACCTGGTGGTCAAGCGCCACGCGCGCGCGCGCGCGCTTCCTCGT	20526
QY	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
DB	20525	CGAATATGAACCGGGTGGCGCGGCGCATCAATGCCCGACACCGCGCGCGCTTCAT	20466
QY	561	GlyGly----ArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg	579
DB	20465	GCCTCGCTGACACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	20406
QY	580	-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlyTyrProGlnThrVal	596

Db 20405 CCACATCCGAGCGCGCATGATTGAGCTATACACATCATGTCAGCAGCAAGTGTGCTTC 20346
Qy 597 ThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThr 616
Db 20345 CCCATGGAACTTGGCTCT-----GAGCTTGAGGAGTCGTCACCGAGTTGCTA 20295
Qy 617 ValSerLeuAsnGlnAspLeuThr-----LeuGlnProTyrGlnValMet 631
Db 20294 AGAGGGTTGATTAGACTTGCATCTGACCTCCATGAACCTGGCCCTCTGCTTATGATGCTT 20235
Qy 632 TrpLeu 633
Db 20234 TGGCTA 20229
RESULT 10
AA559613
ID AA559613 standard; DNA; 22934 BP.
XX AA559613;
AC AA559613;
DT 13-FEB-2002 (first entry)
XX Propionibacterium acnes immunogenic protein encoding DNA #108.
DE SAPHO syndrome; synovitis; acne; pustulosis; hypotosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.
XX Propionibacterium acnes.
OS WO200181581-A2.
XX
PN 01-NOV-2001.
PD
PF 20-APR-2001; 2001WO-US12865.
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX (CORI-) CORIXA CORP.
PA
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
DR
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Claim 1: SEQ ID No 108; 1069pp; English.
XX
XX Sequences AA559506-AA559804 represent DNA molecules encoding
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by P. acnes. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypotosis and
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
CC in infections of bone, joints and the central nervous system, however it
CC is particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acnes in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for P. acnes proteins. These antibodies can be
CC used to downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AAU60884-AAU61133 and AAU67652-AAU67654.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 22934 BP; 4591 A; 7273 C; 6479 G; 4591 T; 0 other;

Alignment Scores:

Pred. No.:	6,15e-47	Length:	22934
Score:	560.50	Matches:	162
Percent Similarity:	43.64%	Conservative:	85
Best Local Similarity:	28.62%	Mismatches:	220
Query Match:	16.45%	Indels:	99
DB:	23	Gaps:	22

US-09-843-007a-2 (1-636) x AA559613 (1-22934)

Qy	118	GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr	137
Db	14744	GGCGACTTTAAAGCGCTGACGGGAAACTCGATTACTTACATGGCTCGGGGGGATTC	14803
Qy	138	LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAval	157
Db	14804	CTATGGTTACACCCCTTCTATGACTCTCCT-----CTCCATGACGAGGTACG	14857
Qy	158	SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValle	177
Db	14858	CGTGACTATCGCTGGATCCGTGAGGAGCTGGGACCACTTGAAGGTTTCCTT	14917
Qy	178	AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer	197
Db	14918	GACGAGCCCATGACCGTGCCCTTGGCTCATCATCGACTTCGTATGAACACACCTCG	14977
Qy	198	AsnGluHisGluTrpAlaGlnArgCysAlaAlaGly---AspProLeuPheAspAsp	216
Db	14978	GATTCCCATCGTGTTCCAGTCTCAGCGCGGATCCGATGCTCTTACGGTAAC---	15034
Qy	217	TyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGlu	236
Db	15035	TACTACGTATGTCGGAC-----ACTGACGAGCCCTACTCCGATGCTCGTATCATC	15085
Qy	237	PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrp	255
Db	15086	TTCTGCGAT-----ACCGAGGACTCGAATTGGTCTCTGGAT	15121
Qy	256	ThrThrPheAsnSerPheGlnTrp-----AspLeuAsnTyr	267
Db	15122	TCCAGCGCTAAGCAGATTCTACTGGCACCGGATTTTCCATCCACCGCATCTCACTT	15181
Qy	268	SerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeu	287
Db	15182	GAAGAGCTCGAGTTATGGAGGAATGCTCGAGCCGCTCGCTTCTGGATGGACTGGGT	15241
Qy	288	ValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCys	307
Db	15242	ATTGACGGATTCCGACTCGACCGCGCTCCGTACTCATTTGAGCGCGGCGGACCACTGC	15301
Qy	308	GluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAla	327
Db	15302	GAGAACTTCCAGCAACTCACAAGATCTCAAGCAGCTGCGCGCCATGGTCGATGAAGAA	15361
Qy	328	AlaProAlaValPhePheLysSerGluAlaIleValHisProAspGlnValGlnTyr	347
Db	15362	TACCAGCGCCGATCTCTGCTGTGTGAAGCTAACCCAGTGGCGCTGACGCTCGTGAT	15421
Qy	348	IleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrp	366
Db	15422	TTTGCAACGGTGAGGAATGCCAGATGGCTTCCATTTCCGTCATGCCCGCTGTAC	15481
Qy	367	AsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHis	386
Db	15482	ATGGGACTCGCTCGGGTTACCGGGAATGCATAGCAGAGATCTCGGTCCACACTCCAC	15541
Qy	387	LeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThr	406

Db 15542 ATCCGGAGCGCTGCAGTGGGGACCTTCTCTGGCAACACGAGAACTCACCTCGAA 15601
Qy 407 Phe---AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425
Db 15602 ATGGTGACCGAAGAGATCGTCACTACATGTGG----- 15634
Qy 426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445
Db 15635 -----GAGGAA 15640
Qy 446 TyrAsnPro---SerThrGlyAspCysArgValSerGlyThrAlaAlaAlaLeuValGly 464
Db 15641 TATGCTCTGTGATCCGGTGGGGTCAACTG---GCATTGCTGGCGCTTGAGCCCT 15697
Qy 465 LeuAlaGlnAspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeu 484
Db 15698 CTTGTCGATACGACGAT-----CCGAGATCCGGCTACTTAACGCCATGTGCTG 15748
Qy 485 SerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAsp 504
Db 15749 TCCCTACCGGCTCTCCGGTGTGTATTACGACGAGAGATTGC---ATGGGGACGAT 15805
Qy 505 AspTrpSerGlnAspSerAsn-----LysSerAspAspSerArgTrp 518
Db 15806 CCTTGGCTGCTGACCGTGTGGGTGGGTACCCCGATGCGATGGGATGACTCGGAGACG 15865
Qy 519 Ala----- 519
Db 15866 GCTGGCTTTTCCACCTGCTCCAGAGGACTTTCATCTACCTCTCATCCGACCTTGGC 15925
Qy 520 HisArgProArgTyr---AsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAla 538
Db 15926 CACGATCCAGACGACGTCAATGTGCGC-----CGCCAGATGGACGCCGTCTCTCC--- 15976
Qy 539 AlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg 558
Db 15977 -----TTGCTGGTGTGGAGCGGTGCCATGCTCGGTATCCCGCCGACACCCCGTC 16027
Qy 559 PheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleGlyTyrIle 578
Db 16028 TTGGGACCGGCGGACTTCCCGATAGCGGGCGGACATGGCGGTGATGCTTCCTG 16087
Qy 579 ArgAsnAsn-----AlaLeuLeuAlaPheGlyAsnPheSerGlyTyrProGlnThr 595
Db 16088 CGTCAACAGCAGCAGAAACGGTCTGCTGGCTGCTAACTTCTCCGATCTGAGCGGATG 16147
Qy 596 ValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyLys 615
Db 16148 GTTGCCTTTCATCTCCCAANTCCGGGCATGACGGGCTCTTCTCTCATCTCATGTCAG 16207
Qy 616 ThrValSer---LeuAsnGlnAspLeuThrLeuGln-----ProTyrGlnVal 630
Db 16208 GACGGCAACCACTAAAGCTGACGGAACACTGTCCGTACCGTTGTGGCCATATGGCTAT 16267
Qy 631 MetTrpLeuGluIleAla 636
Db 16268 CGATGGCTGCAAGTGTCC 16285

RESULT 11
ID AAS54127
ID AAS54127 standard; DNA; 3303 BP.
XX AAS54127;
XX AAS54127;
DT 13-FEB-2002 (first entry)
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #258.
XX Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX Pseudomonas aeruginosa.

XX W0200170955-A2.
XX 27-SBP-2001.
XX 21-MAR-2001; 2001WO-US09180.
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207272P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITEA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX P-PSDB; AAU36268.
XX WPI; 2001-611495/70.
XX P-PSDB; AAU36268.
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX Claim 27; Seq ID No 7764; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins,
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 3303 BP; 578 A; 1171 C; 1084 G; 470 T; 0 other;

Alignment Scores:
Pred. No.: 3,54e-46 Length: 3303
Score: 541.50 Matches: 165
Percent Similarity: 41.60% Conservative: 90
Best Local Similarity: 26.92% Mismatches: 223
Query Match: 15.89% Indels: 135
DB: 23 Gaps: 25
US-09-843-007A-2 (1-636) x AAS54127 (1-3303)

Qy 96 AsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyValCysTyr----- 112
Db 31 AACGACCGCAATGGTACAGGAC-----GGGTGATCTACAGGTCCAC 75
Qy 113 ValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLysAsp 125
Db 76 GAGAAATCTTCTACGACGCCAACACGATGGCATCGGACTTCGGCGGCTCATCGAG 135
Qy 126 LysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLys 145
Db 136 AAGCTCGACTACATCGCGGACCTCGCGGTGAGAACACTCTCTGGCTGCTCCGCTTCTAC--- 192
Qy 146 CysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnPro 165

Db 193 ---CGTCCGACGCGCGACGCGCTACGACATCGCCGCTACGCTGGCGTGCACAGC 249
Qy 166 AlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIle 185
Db 250 GACTACGCGACGCTCGCGCGCGCGCTGTCATCGCGCGCGCGCTGCTG 309
Qy 186 SerAlaValAlaAspPheIlePheAsnHisThrSerAsnGluHisGluTrp 202
Db 310 CGGTGATATACGAGCTGGTATCAACCATCTCCGACCGACATCCCTGGTTATCCGT 369
Qy 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222
Db 370 GCCCGCCACGCGAAGAGGATCGCGCGCGCGAC-----TACTACGTCTGGTTCGAC 423
Qy 223 ArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisPro 242
Db 424 -----ACGACGAGAAATACACAGGTACG---CGGATCATCTTCATCGAC----- 465
Qy 243 GlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThr----- 256
Db 466 -----ACGAGCAGTCCCACTGGACCTGGACCGCGTAGTACCCACAGTAC 510
Qy 257 -----ThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPhe 273
Db 511 TACTGGCACCCTTCTATTCCACGACGCGGACCTGAACCTTCGACACCCGCGAGTCTGT 570
Qy 274 ArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMet 293
Db 571 CGCGAGGTCTCGGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 630
Qy 294 AspAlaValAlaPheIleTrpIleGlnMetGlyThrSerCysGluAsnLeuProGlnAla 313
Db 631 GACGCGATCTCTACCTGATCGAAGCGACGCGACCGACGACGCGAGTCTGCGGAGACC 690
Qy 314 HisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePhe 333
Db 691 CACCAAGTCTCAAGCGCATCCGCGCGAGCTGGACGCGCATCTCCGACCGCATGCTG 750
Qy 334 LysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly----- 349
Db 751 CTGGCGAGGCGCAATCAGTGGCGGAGACAGACCGCGCGTACTTCCGCGCGGAGATGGC 810
Qy 350 -----GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuTrpAsn 367
Db 811 GCGAGGGCGAGCAATGCGACATGGCTTCCACTTCCCGCTGATCGCGGCTGATGATG 870
Qy 368 ThrLeuAlaThrArgGluValAlaAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeu 387
Db 871 GCATCGCCACGAGGATCGCTATCCGATCCGACATCTCGCGCAGACCCGCGACATC 930
Qy 388 ProGluHisThrAlaTrpValAsnTyrValArgSerHisAspIleGlyTrpThrPhe 407
Db 931 CCGGCCAATTGCAATGGCGGATCTTCTCGCGCAACCCACGACGCGTGCACCTGGAGATG 990
Qy 408 AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn 427
Db 991 GTCACCGAC-----GACGAGCGGACTATCTC----- 1017
Qy 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
Db 1018 -----TGGAC 1023
Qy 448 ProSerThrGlyAspCysArgValSer-----GlyThrAlaAlaAlaLeuValGly 464
Db 1024 CACTATCGCGCGACCGCGCGCGCTCAACCTGGGCTATCCGCGCGCTGCGCGCG 1083
Qy 465 LeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeu 484
Db 1084 CTGGTGGAGGTGACCGCGCG-----CGCATCGAGCTGTCACAGCTGCTGCTG 1134
Qy 485 SerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr----- 500

Db 1135 TCGATCGCGGCGACGCGACCTGTACTACGCGCGAGAGATCGCATGGCGGACACATC 1194
Qy 501 ---LeuAsnAspAsp-----TrpSerGlnAspSerAsn--- 511
Db 1195 TACTCTCGCGACCGCGACCGCGCGCGCGATCGATCGATGGTGGTGGACCGCACGCG 1254
Qy 512 -----LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu 528
Db 1255 GGCTTCTCCCGCGCGCGCGCGCGAGCTGGTG---CTGCCGCGGATTCGACCCGCTG 1311
Qy 529 Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly 540
Db 1312 TACGGCTACACGACGATCAACGTCGAGCGCGCGCGCGCGCGATTCG----- 1362
Qy 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
Db 1363 ---CTGCTCACTGAGTGGCGCGCTCTCGCGTGGCGACCGCAGAGGCGCTTCGCG 1419
Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnLysHisIleIleGlyTyrIle----- 578
Db 1420 CGCGCGACCTGAAGATGCTCGCGCGCGCGCGCGCGCGCGCGCTACCTCGCGCGAA 1479
Qy 579 -----ArgAsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrPro 593
Db 1480 TACGCCGAGCGCGACGCGCGAGGACGATCTCTGCTGGCGCACTGTGCGCGCGCG 1539
Qy 594 GlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis----- 609
Db 1540 CAGCGCGTG-----GAACTGGACCTCGCGCGCGCGCGCGCGCGCGAGGTG 1581
Qy 610 -----AspLeuIleGlyGlyLysThrVal-----SerLeuAsnGlnAspLeu 623
Db 1582 CCGGTGAGATGATGCGCGCGATGCTTCCCGCGCGATCGCGCGCGTACTTACTCTG 1641
Qy 624 ThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
Db 1642 ACCCTCGCGCTACGCTTCTACTGTTCTTACTCGCGCG 1680

RESULT 12
AAH67492
ID AAH67492 standard; DNA; 1794 BP.
XX
AC AAH67492;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2527.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PP 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR P-PSDB; AAG92273.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing

expression profile or pattern of a gene and identifying homologous gene

Claim 8, SEQ ID NO: 2527; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium glutamicum. These sequences are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Corynebacterium, and identifying a homolog of a gene derived from corynebacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 1794 BP; 425 A; 542 C; 432 G; 395 T; 0 other;

Alignment Scores:

Pred. No.: 2,64e-45 Length: 1794
Score: 529.50 Matches: 153
Percent Similarity: 46.32% Conservative: 105
Best Local Similarity: 27.47% Mismatches: 218
Query Match: 15.54% Indels: 81
DB: 22 Gaps: 22

US-09-843-007a-2 (1-636) x AAH67492 (1-1794)

118 GlyAspLeuLysGlyLeuLysAspLysLeuLysPheGlnGluLeuGlyLeuThrTyr 137
123 GGATCGTTGAAGGCTGACCGAAGAACTGGATTACATCCAGTGGCTCGCGTGGATTGC 282
138 LeuHisLeuMetProLeuPheLysCysProGluLysSerAspGlyTyrAlaVal 157
283 ATTTGGATCCCATCGTTTATGATCCCA-----CTGCGCGGACGCGGTACGATC 336
158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValile 177
337 CGCAACTTCGCTGAATCTGCGCGAATTCGGACCGCTCGATGACTTCGTGGAACTCGTT 396
178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheLeuPheAsnHisThrSer 197
397 GACCACGCCACCGCCCTGCGCTGCTGTTATCACCGACTTGTGTCATGATCAACCTCC 456
198 AsnGluHisGluTyrAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215
457 GACCAGACCGCATGGTTCCAGAA---TCCCGCGCGACCCACCGCCCTTCAGCGAT 513
216 PheTyrThrPheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGlu 235
514 TTCTATGTGTGGAGCGATGAT-----CCACCCCTGTACAAACGAGCC--CGCATC 561
236 IlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTyrVal 254
562 ATCTTGTAGATACAGAAAGATCCACTGACCTATGATCCGTTGCGTGGCTGCTACTTC 621
255 TrpThrThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArg 274
622 TGGCACCCTTCTCTCCCAACCAAGACCTCACTACGACCAACCCCGCGAGTCCAGAG 681
275 AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp 294
682 GCATGCTAGATGCTTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 741
295 AlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis 314
742 GCGCTTCTTATCTTTTGAACGCGAGGACCAACGCGGAAACCTCAAGAAACCCAC 801
315 AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys 334

802 GATTTCCTCAAACTGTGTGCTCTGTCTATTGAGAGGATACCCCGCGCATCTGCTCTC 861
335 SerGluAlaIleValHisProAspGlnValValGlnTyrIleGlyGln----- 350
862 CGAGAACCCAAACCAATGGCCCAAGATGTGTGCTGCAATACTTCTGCTGAAAGAACAAAGGC 921
351 AspGluCysGlnIleGlyTyrAsnProLeuGluMetAlaLeuLeuTyrAsnThrLeuAla 370
922 GATGATGCGCATGGCTTCCACTTCCCTTTCATGCGCGCATCTTCATGGAGTTCGC 981
371 ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
982 CRAAGTTCACGACCCCGCATGAGATCTCTGCGCAACCCCGGAGATTCACCAAGACT 1041
391 ThrAlaTyrValAsnTyrValArgSerHisAspAspIleGlyTyrThrPhe---AlaAsp 409
1042 GCCCAATGGGTATTTCTGCTGATATCATGATGAGTACCTTCAATGGTGTCTCGAT 1101
410 GluAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn----- 427
1102 GAGGAACGCGAGTACATG-----TACTCCCAATTGCGCTCCGAACCT 1143
428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
1144 CGCATGCGCGCCAAC----- 1158
448 ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467
1159 -----GTAGGATCCGCGAGCGCTTCCCACTGCTGAA 1194
468 AspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeuSerThrGly 487
1195 GCGCAACGCG-----AACCACTGGAAGTCTTCAAGTGTGTGCTGCTTCTACCT 1245
488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
1246 GGTCTACCGGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1305
503 AspAspAsp-----TrpSerGlnAspSerAsn----- 511
1306 GACCGCGACCGAGTGGCGACCCCATCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
512 LysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeu---TyrAla 530
1366 AAAGCTGATCTGCAAGCGCTTACCTTCCAGCGATCCAAATGATCAATACCGCTACGCC 1425
531 GlnArgAsnAspProSer-----ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis 548
1426 CAAATTAACGTGGAAGCAACCTCAACCGCGAAACTCCCTGCTGCTGCTGCTGCTGCTGCT 1485
549 MetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsn 568
1486 CAAATCTTATCCGCAAGCAGTACCGCGCATTTGCTGCGGAACTTACCGTGAAGTGTCC 1545
569 ThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsn-----AlaLeuLeuAla 585
1546 TCCACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1605
586 PheGlyAsnPheSerGlyTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro 605
1606 GTCAACCAACATGACCAATATCTCAGGCGATC-----TCGCTTGAATTCGCTGAA 1656
606 PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu 619
1657 TTTGACGAGCAACCCCTCGAGAGATGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1716
620 AsnGln---AspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIle 635
1717 CGGAGTGGATGTCTACCTTACCGCTTACCGGATTTCTCTGCTGCTGCTGCTGCTGCTGCT 1767

RESULT 13
ABS65343
ID ABS65343 standard; DNA; 1981 BP.

XX AC ABS55343;
 XX DT 15-NOV-2002 (first entry)
 XX DE DNA encoding C. glutamicum metabolic pathway (MP) protein #2.
 XX KW Metabolic pathway protein; MP; cell metabolism; amino acid; vitamin;
 XX KW cofactor; nucleoside; nucleoside; trehalose; fine chemical production;
 XX KW organic acid; non-proteinogenic amino acid; purine base; carbohydrate;
 XX KW pyrimidine base; lipid; unsaturated fatty acid; diols; polyketide;
 XX KW aromatic compound; food industry; animal feed; cosmetic industry;
 XX KW pharmaceutical industry; gene; ds.
 XX OS Corynebacterium glutamicum ATCC 13032.
 XX PN W0200251231-A1.
 XX PD 04-JUL-2002.
 XX PF 22-DEC-2000; 2000WO-EPI13143.
 XX PR 22-DEC-2000; 2000WO-EPI13143.
 XX PA (BADI) BASF AG.
 XX PI Pompejus M, Kroeger B, Zelder O, Schroeder H;
 XX DR NFI; 2002-643289/69.
 XX DR P-PSDB; A3G80322.
 XX PT New metabolic pathway genes of Corynebacterium glutamicum for producing
 XX PT fine chemicals, e.g. lipids, (un)saturated fatty acids, vitamins,
 XX PT cofactors or enzymes used in food, feed, cosmetics or pharmaceutical
 XX PT industries -
 XX PS Claim 1; Page 95-98; 176pp; English.
 XX CC The present invention relates to the isolation of Corynebacterium
 XX CC glutamicum metabolic pathway (MP) proteins, and the polynucleotide
 XX CC sequences encoding them. The MP proteins are enzymes involved in
 XX CC the metabolism of molecules important for the normal functioning
 XX CC of cells (e.g. amino acids, vitamins, cofactors, nucleosides and
 XX CC nucleosides, or trehalose). The polynucleotide sequences encoding
 XX CC the MP proteins are useful for producing fine chemicals, particularly
 XX CC organic acids, non-proteinogenic amino acids, purine and pyrimidine
 XX CC bases, nucleosides, nucleotides, lipids, (un)saturated fatty acids,
 XX CC diols, carbohydrates, aromatic compounds, vitamins, cofactors,
 XX CC polyketides and enzymes. The fine chemicals are useful in the food,
 XX CC animal feed, cosmetic or pharmaceutical industries. ABS55342-ABS55364
 XX CC encode the C. glutamicum MP proteins of the invention.
 XX SQ Sequence 1981 BP; 470 A; 596 C; 476 G; 439 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.05e-45 Length: 1981
 Score: 529.50 Matches: 153
 Percent Similarity: 46.32% Conservative: 105
 Best Local Similarity: 27.47% Mismatches: 218
 Query Match: 15.54% Indels: 81
 DB: 24 Gaps: 22
 US-09-843-007A-2 (1-636) x ABS55343 (1-1981)
 QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
 DB 380 GGATCGTTGAAGGCGCTGACCGGAAACTGGATTACATCCAGTGGCTCGCGTGGATTGC 439
 QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyLysTyrAlaVal 157
 DB 440 ATTGGATCCACCGCTTTTATGATTCCCA-----CTGCGCGACGCGGTACGATATC 493
 QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
 DB 494 CGCAACTTCGGTGAATCCTCGCGAATTCGGCAGCTCGATGACTTCGTGAACTCGTT 553
 QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197
 DB 554 GACCAAGCCACCGCGCTGCTGTATACCGGACTTGGTCAATGATCACACTCC 613
 QY 198 AsnGluHisGluTTPAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215
 DB 614 GACCAAGCCACCGCTGCTGTATACCGGACTTGGTCAATGATCACACTCC 670
 QY 216 PheTyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu 235
 DB 671 TTTATGTGTGAGCGATGAT-----CCACCCCTGTACACGAGCC---CGCATC 718
 QY 236 IlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTTPVal 254
 DB 719 ATCTTTGTAGATACAGAAAGATCCAACTGGACCTATGATCCGGTGGCGGAGTACTTC 778
 QY 255 TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTTPValPheArg 274
 DB 779 TGGCACCGCTTCTCTCCACCCACGACCTCACTACGACACACCCCGCAGTCCAGAG 838
 QY 275 AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp 294
 DB 839 GCCATGCTAGATGCTCTGCGTTCTGGCTGGACCTGGAGCTTGATGGTTTCGAGTAGAT 898
 QY 295 AlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis 314
 DB 899 GCCGTTCTCTATCTTTTGAACGCGAGGACCAACGCGGAAACCTCAAGAAACCCAC 958
 QY 315 AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys 334
 DB 959 GATTTCTCAAACTGTGCTCTCTCATTTGAGAGGAATACCCCGCGGAGTCTCTGCTC 1018
 QY 335 SerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGln----- 350
 DB 1019 GCAGAGGCAACCAATGCGCCCAAGATGTGTGCGAATCTTCGTGAAACACAAAGGC 1078
 QY 351 AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAla 1078
 DB 1079 GATGAATGCCATGCGCTTCCACTTCCCTTTGATGCGCGCATCTTCCATGCGAGTTCG 1138
 QY 371 ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
 DB 1139 CAAGTTCCAGCACCCCGATCAGTGAATCTGCGCAACACCCCGGAGATCCCAAGACT 1198
 QY 391 ThrAlaTTPValAsnTyrValArgSerHisAspAspIleGlyTyrThrPhe---AlaAsp 409
 DB 1199 GCCCATGGGTATTTCTCGGTATATCATGATGAGCTCACCTTGAATGTCTCTCCGAT 1258
 QY 410 GluAspAlaAlaTyrLeuGlyLysSerGlyTyrAspHisArgGlnPheLeuAsn----- 427
 DB 1259 GAGGAACGCGACTACATG-----TACTCCCAATTCGCTCCGAACT 1300
 QY 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
 DB 1301 CGCATGCGCGCCAC----- 1315
 QY 448 ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467
 DB 1316 -----CTAGCAATCCGAGCGCTTTTCCCACTGCTTGA 1351
 QY 468 AspAspProHisAlaValAspArgIleLysLeuTyrSerIleAlaLeuSerThrGly 487
 DB 1352 GCGCAGCCG-----AACGAGTGAACCTCTTACGGTTTGTGTCTCTACTCT 1402
 QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
 DB 1403 GGCCTACCGTGTGATTACGTCATGAATTCGATGGCGGACATATCTGGCTCCAC 1462
 QY 503 AspAspAsp-----TrpSerGlnAspSerAsn----- 511

QY 371 ThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
Db 339094 CRAAGTTCACGCCCGATCAGTGATCTCGGCCAACACCCCGAGATCCCAAGACT 339153
QY 391 ThrAlaTTPValAsnTyrValArgSerHisAspAlleclYTrpThrPhe---AlaAsp 409
Db 339154 GCCCAATGGGTATTTCTCGGTATATCATGATGACCTCACCTTGAATGGTCTCCGAT 339213
QY 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn---- 427
Db 339214 GAGGACACGAGCTACATG-----TATCCCAATCCCTCCGAACCT 339255
QY 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgValProPheGlnTyrAsn 447
Db 339256 CGCATCGCGCCCAAC----- 339270
QY 448 ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467
Db 339271 -----GTAGGAATCCGCGAGCGCTTTCCCACTGTTGTAA 339306
QY 468 AspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly 487
Db 339307 GCGCAGCGC-----AACCACTGGAATCTCTTCACGGTTGTGTCTCTACCT 339357
QY 488 GlyLeuProLeuLeuTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
Db 339358 GGTCACTCCCTGTTGTTATTACGGTGATGAATTTGGCATGGCGACAAATATCTGGCTCCAC 339417
QY 503 AspAspAsp-----TTPSerGlnAspSerAsn----- 511
Db 339418 GACCGCAGGAGTGCACCCCGCATGTCATGCTCCACGCGCAACGGTGGTCTCTCC 339477
QY 512 LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu---TyrAla 530
Db 339478 AAAGCTGATCTGGAAGCGCTGACCTTCCAGCGATCCAAATGATCAATACGGCTACGCC 339537
QY 531 GlnArgAsnAspProSer-----ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis 548
Db 339538 CAGTAAACCTGGAAGACCACTCAACCGGAAACTCCCTGCTGGCTGCTCCGAAC 339597
QY 549 MetIleAlaValArgGlnSerAsnProArgPheAspGlyArgLeuValThrPheAsn 568
Db 339598 CAATCTTATCCGCAAGTAGTACCGCGCATTTGGTGGCGGAACCTACCGTGAAGTCTCC 339657
QY 569 ThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsn-----AlaLeuLeuAla 585
Db 339658 TCCACCAATGAGTCAGTGTGACATTTTACGAGAACACAAAGGCGCAACCACTTTGTGT 339717
QY 586 PheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro 605
Db 339718 GTCAACACATGAGCAATATCTCAGCGATC-----TCGCTTGTATTGCTGAA 339768
QY 606 PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu 619
Db 339769 TTTGCAAGACACACCCCTCGAGAGATGTCGGCGGCGCATGTTCCCTACCACTTGTGAA 339828
QY 620 AsnGln---AspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIle 635
Db 339829 CGGAGTGGATTCATCTTAGCCCTCACCAGATCTTCTGGTTGATCTC 339879
RESULT 15
AAH68532
ID AAH68532 standard; DNA; 349980 BP.
XX
AC AAH68532;
XX
XX
DT 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 7067.
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX
XX Disclosure; SEQ ID NO: 7067; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids.
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
SQ Sequence 349980 BP; 80900 A; 98397 C; 92139 G; 78544 T; 0 other;

Alignment Scores:
Pred. No.: 6,14e-42 Length: 349980
Score: 529.50 Matches: 153
Percent Similarity: 45.32% Conservative: 105
Best Local Similarity: 27.47% Mismatches: 218
Query Match: 15.54% Indels: 81
DB: 22 Gaps: 22

US-09-843-007A-2 (1-636) x AAH68532 (1-349980)
QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
Db 38335 GGATCGTTGAAGGCTGACCGAATAATGATATACATCCAGTGGCTCGGGGTGATTCG 38394
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
Db 38395 ATTTGGATCCACCGCTTTTATGATTCCTCCCA-----CTGCGCGACGGGGTACGATATC 38448
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 38449 CGCAACTTCGGTGAATCTCTCCGAATTCGCGACCGCTGATGATCTTGTGGAATCGTT 38508
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
Db 38509 GACCACGCCACCGCGGTGGCTGTGTTATCCGACTTGTGTCATGATATCACACCTCC 38568
QY 198 AsnGluHisGluTTPAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215
Db 38569 GACCACGACGATGTTTCCAGAA---TCCCGCGCGACCCCAACCGGCCCTTACGGAGAT 38625

```
QY 216 PheTyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu 235
DB 38626 TTCTATGTGTGGAGCGATGAT-----CCACCGCTGTACAGAGGCC---GCGATC 38673
QY 236 IlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTrpVal 254
DB 38674 ATCTTTGTAGATACAGAGAAATCCAACTGGACCTATGATCCGGTGGCGAGTACTTC 38733
QY 255 TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArg 274
DB 38734 TGGCACCGCTTCTTCTCCCAACACAGACCTCAACTACAGACACCCCGCAGTCCCAAG 38793
QY 275 AlaMetAlaGlyMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp 294
DB 38794 GCGATGTAGATGCTTGGCTTCTGGCTGGACCTTGGACCTTGATGGTTCCGAGTAGAT 38853
QY 295 AlaValAlaPheIleTrpPheGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis 314
DB 38854 GCGGTTCTCTATCTTTTGAACGCGAGGCCAACCAACGCGGAAAACCTCAAGAGAAACCCAC 38913
QY 315 AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys 334
DB 38914 GATTTCCCTCAAACTGTGTGCTCTGTATGTAGAGAGAAATACCCCGCGCGAATCTGCTC 38973
QY 335 SerGluAlaIleValHisProAspGlnValValGlnTyrIleGlyGln----- 350
DB 38974 GCAGAGCCACCAATGCCCCCAAGATGTGTGTAATCTCGGTGAAAAGACACAAAGCC 39033
QY 351 AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeuAla 370
DB 39034 GATGAATCCACATCGCCCTTCCACTTCCCTTTGATGCGCGCGCATCTTTCATGGGAGTTCGC 39093
QY 371 ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
DB 39094 CNAAGTTACAGCCACCCCGATCAGTGAGATCTTGGCCACACCCCGAGATTTCCCAAGACT 39153
QY 391 ThrAlaTrpValAsnTyrValArgSerHisAspIleGlyTrpThrPhe---AlaAsp 409
DB 39154 GCGCAATGGGTATTTCTCGGTAAATCATGATGAGCTCACCCCTGAAATGGTCTCCGAT 39213
QY 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn----- 427
DB 39214 GAGGAACGAGCTACATG-----TACTCCCAATTCGCTCCGAACT 39255
QY 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
DB 39256 CGCATGCGCGCCAAC----- 39270
QY 448 ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467
DB 39271 -----GTAGAAATCCGACGCGCTTTCCCCACTGCTTGAA 39306
QY 468 AspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly 487
DB 39307 GCGGACCGC-----AACCGCTGGAACCTCTTCAAGGTTGTTGCTGTCTTACTCT 39357
QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
DB 39358 GCGTCACCGCTGTTGTTATTACGGTATGAAATGGCATGGCGGCGACATATCTGGCTCCAC 39417
QY 503 AspAspAsp-----TrpSerGlnAspSerAsn----- 511
DB 39418 GACCGGACGGAGTGGCGACCCCATCGATGGTCCACGACCGACCGAGTGGTTCTCC 39477
QY 512 LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu---TyrAla 530
DB 39478 AAAGCTGATCTGGAAGCGCTGTACCTCCAGCGATCCCAAAATGATCAATACGGTACGCC 39537
QY 531 GlnArgAsnAspProSer-----ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis 548
DB 39538 CAAGTAAACGTTGGAAGGCAACTCAACCGCGAAAACTCCCTGCTCGCTGGCTCGGAAC 39597
QY 549 MetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsn 568
```

```
DB 39598 CAAATCCTTATCCCGAAGCAGTAGTACCGCGCATTTTGGTCCCGGAACCTACCGTGAAGTGTC 39657
QY 569 ThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsn-----AlaLeuLeuAla 585
DB 39658 TCCACCATGATGATCAGTGTGACATTTTACGAGAACACAGGCGCCAAACCATTTTGTGT 39717
QY 586 PheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro 605
DB 39718 GTCACCAACATGAGCAAAATATCTCTCAGGCAGTC-----TCGCTTGTATTGCGTGAA 39768
QY 606 PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu 619
DB 39769 TTTCCAGACACACCCCTCCGAGAGATGTCCGGCGGCGAGCTGTTCCCTTACCATTTCTGAA 39828
QY 620 AsnGln---AspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIle 635
DB 39829 CCGGAGTGATGTGCACTTTAGCCCTCACGGATTCTTCTGTTGATCTC 39879
```

Search completed: November 8, 2003, 20:28:51
Job time : 852 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2003, 19:59:38 : Search time 110 Seconds

(without alignments)
2551.997 Million cell updates/sec

Title: US-09-843-007A-2

Perfect score: 3408
Sequence: 1 MLTPTQVGLILQVLTIRL.....VSLNQDLTLPQVYVWLEIA 636

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-D=/cgn2_1/USPTO.spool.p/US09843007/runat_07112003_170420_25382/app_query.fasta_1.775
-DB=Issued Patents NA -QFWT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09843007 @CEN 1 1 103 @runat_07112003_170420_25382 -NCPUP=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3408	100.0	2914	3	US-08-737-752A-1
2	543.5	15.9	3414	4	US-09-252-991A-8258
3	543.5	15.9	4125	4	US-09-252-991A-8065
4	509.5	15.0	4403765	3	US-09-103-840A-2
5	509.5	15.0	4411529	3	US-09-103-840A-1
6	504	14.8	1704	1	US-08-103-840A-2
7	504	14.8	1704	1	US-08-528-199-5
8	488	14.3	2889	1	US-08-537-002A-4
9	488	14.3	2889	3	US-08-863-010-4
10	488	14.3	2889	3	US-09-024-429-4
11	488	14.3	3600	1	US-08-537-002A-5
12	488	14.3	3600	3	US-08-863-010-5

13	488	14.3	3600	3	US-09-024-429-5
14	404.5	11.9	1728	4	US-09-107-532A-2079
15	317	9.3	1683	4	US-09-134-001C-316
16	314.5	9.2	1635	4	US-09-107-532A-482
17	313.5	9.2	1782	1	US-08-374-155A-13
18	313.5	9.2	1782	2	US-08-785-396-13
19	302.5	8.9	4776	2	US-08-852-401-1
20	301.5	8.8	2074	1	US-08-039-777-2
21	301.5	8.8	2074	1	US-08-611-361A-2
22	287	8.4	4784	4	US-09-634-238-24
23	283.5	8.3	1794	1	US-08-374-155A-11
24	283.5	8.3	1794	2	US-08-785-396-11
25	280	8.2	1704	2	US-08-374-155A-15
26	280	8.2	1704	2	US-08-785-396-15
27	277.5	8.1	3895	4	US-09-961-527-201
28	259	7.6	4105	4	US-09-634-238-182
29	251.5	7.4	1608	3	US-08-137-077-1
30	251.5	7.4	1650	4	US-09-107-532A-3186
31	248	7.3	1668	4	US-09-107-532A-1068
32	245.5	7.2	6749	4	US-08-961-527-84
33	244	7.2	1890	1	US-08-374-155A-1
34	244	7.2	1890	2	US-08-785-396-1
35	243.5	7.1	1803	1	US-08-374-155A-9
36	243.5	7.1	1803	2	US-08-785-396-9
37	234.5	6.9	1305	1	US-08-374-155A-2
38	234.5	6.9	1305	2	US-08-785-396-2
39	233	6.8	2100	1	US-07-927-316A-1
40	227	6.7	1734	1	US-08-551-437-4
41	227	6.7	1734	3	US-09-004-225-4
42	227	6.7	1734	3	US-08-084-346-4
43	227	6.7	1734	3	US-09-104-704-4
44	212	6.2	1692	4	US-09-107-532A-2582
45	212	6.2	1692	4	US-09-107-532A-2583

ALIGNMENTS

RESULT 1
US-08-737-752A-1
; Sequence 1, Application US/08737752A
; Patent No. 6265635

GENERAL INFORMATION:

APPLICANT: Kossman, Jens
APPLICANT: Kossman, Jens
APPLICANT: Welsh, Thomas

TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE
OF FACILITATING THE SYNTHESIS OF LINEAR
ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND
MICROORGANISMS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08737,752A

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: DE P 44 17 879.4

FILING DATE: 18-MAY-1994

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: DE P 44 47 388.5

FILING DATE: 22-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: James F. Haley, Jr., Esq.
REGISTRATION NUMBER: 27,794
REFERENCES/DOCKET NUMBER: GFB-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2914 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Neisseria polysacchara
IMMEDIATE SOURCE:
LIBRARY: genomic library in pBluescriptII SK
CLONE: pNB2
FEATURE:
NAME/KEY: CDS
LOCATION: 957..2867
US-08-737-752A-1

Alignment Scores:

Pred. No.: 0 Length: 2914
Score: 3408.00 Matches: 636
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-843-007A-2 (1-636) x US-08-737-752A-1 (1-2914)

QY	1	MetLeuThrProThrGlnGlnValGlyLeuLeuLeuGlnTyrLeuLysThrArgIleLeu	20
DB	957	ATGTTGACCCCAACGAGGAGTCGGTTGATTTTACGTACTCAAAACAGCATCTTG	1016
QY	21	AspIleTyrThrProGlnGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe	40
DB	1017	GACATCTACACGCGCGAACAAGCGCGCGCATCGAAATTCGAAAGACTGGCGGCGAGTTT	1076
QY	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
DB	1077	TCGCGCGCGATGATACGCAATTTCCCAACACTGATGACGACTCGACAGCGTGATCGGC	1136
QY	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr	80
DB	1137	AACACGAGCCCTGCTGCTATGCTGGAATGCTGCTGCGCAGCATGGCAAGCTAT	1196
QY	81	SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnProAspTrp	100
DB	1197	TCCCAACGCAACTCATCTTAAAGATATCGATCGCGCGGAAACCAACCCCGATTGG	1256
QY	101	IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
DB	1257	ATTGTGTCACCAACAAACAGTCGCGGCGTGTGTGCTGATTTGTTGCGCGGATTG	1316
QY	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrIleuHisLeu	140
DB	1317	AAGGCGTGAAGATAAATTCCTATTTTCAAGAGCTTCGTTGACTTATCTGCACCTG	1376
QY	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr	160
DB	1377	ATGCCGCTGTTAATAGCCCTGAAGGCAAAAGCGCGCGCTATGCGGTCCACAGCTAC	1436
QY	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu	180
DB	1437	CGCGATGCAATCGGCACTGGGCACATAGAGGAGCTTGGCGAAAGTCAATTGTGCGCTG	1496
QY	181	HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis	200

DB	1497	CACGAAGCCGGCATTTCCGGCGTGTGCGATTTTATCTTCAACACACACCTCCACGAACAC	1556
QY	201	GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220
DB	1557	GAATGGGCGCAACGCTGCGCGCGCGCGCGCTTTTTCGACAATTTCTACTATAATTTC	1616
QY	221	ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240
DB	1617	CCCGACCGCGGATGCCGCAATACGACCGCAACCTGCGGCAATCTTCCCGGACGAG	1676
QY	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrThrPheAsnSer	260
DB	1677	CACCGCGCGGCTTCTCGCAACTCGAAGACGCGCTGGGTGTGCGACACCTTCANTCC	1736
QY	261	PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet	280
DB	1737	TTCCAATGGGACTTGAATTACAGCAACCCGCGGTATTTCCGCGCAATGGCGGCGAAATG	1796
QY	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp	300
DB	1797	CTGTTCCCTTGCCTTGGCAACTTGGCGGTGACATCTCGGTATGGAATGGGTGCTTTATTGG	1856
QY	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320
DB	1857	AAACAAATGGGACAGCTCGCAAAACCTGCGCAGCGCGACGCCCTCATCGCGCGTTC	1916
QY	321	AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340
DB	1917	AATGCCGTATGCGTATTTGCCCGCGCGCGCGGTCTTCTCAAAATCGAAGCATGTCAC	1976
QY	341	ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360
DB	1977	CCGACCAAGTCTGTCATACATCGGCGAGGACGATCGCAATCGGTTTACCAACCCCTG	2036
QY	361	GlnMetAlaLeuLeuTrpAsnThrIleuAlaThrArgGluValAsnLeuLeuHisGlnAla	380
DB	2037	CAAAATGGCATTTGTTGGAAACACCTTTGCCACGCGCGGAGTCAACCTGCTCCATCAGCG	2096
QY	381	LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis	400
DB	2097	CTGACTACCCGCAACACTGCGCGAGCATACCGCTGGGTCTCACTAGCTCCGACGCCAC	2156
QY	401	AspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr	420
DB	2157	GACGACATCGGCTGCGCTGTCGATGACAGCGCGCATATCTGGGGGATAAGCGCGCTAC	2216
QY	421	AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg	440
DB	2217	GACCAACCGCAATTTCTTCAACCGCTTCTTGTCAACCGCTTTTCGACGCGAGCTTCGCTGT	2276
QY	441	GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla	460
DB	2277	GGCGTACCGTTCCATACAACCAAGCACAGCGGACTGCGGTGTCAGTGGTACAGCGCGG	2336
QY	461	AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr	480
DB	2337	GCAATGGTGGGCTTGGCGCAAGACGATCCCAACCGCTTGACCCGCTCAAACTCTGTAC	2396
QY	481	SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500
DB	2397	AGCATTTGCTTTGAGTACCGGCGGTCTGCGGTGATTTTACCTAGGCGACGAACTGGGTAG	2456
QY	501	LeuAsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis	520
DB	2457	CTCAATGACGACGACTGGTTCGCAACAGCAATTAAGACGACGACGCGGTGGGCGCAC	2516
QY	521	ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540
DB	2517	CGTCCGCGCTACACGACGCGCTGTACGCGCAACGACGACGATCCGTCGACCGCAGCGGG	2576
QY	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
DB	2577	CAAACTCATCAGGCTTGGCGCATATGATTTGCGCGTCCGCAAAAGCAATCCGCGCTTCGAC	2636

QY 561 GlyArgGluValThrPheAsnThrAsnAspLysHisIleLeuGlyTyrIleArgAsn 580
DB 2637 GCGGAGGCTGGTACATTCACCAACCAACAGCAATCGCTACATCCGCAAC 2696
QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
DB 2697 AATGCGCTTTTGGCATTCGGTAATTCAGCAATATCGCAAAACGTTACCGCGCATACC 2756
QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
DB 2757 CTGCAAGCCATGCCCTTCAAGCGGACGACCTTCATCGGTGGCAAACTGTGACGCTGAAT 2816
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTyrPheLeuGluIleAla 636
DB 2817 CAGGATTGACGCTTCAGCCCTATCAGGTTCATGTGCTCGAATCGCC 2864
RESULT 2
US-09-252-991A-8258
; Sequence 8258, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8258
; LENGTH: 3414
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8258
Alignment Scores:
Pred. No.: 4,01e-55 Length: 3414
Score: 543.50 Matches: 169
Percent Similarity: 41.59% Conservative: 93
Best local Similarity: 26.83% Mismatches: 227
Query Match: 15.95% Indels: 141
DB: 4 Gaps: 26
US-09-843-007A-2 (1-636) x US-09-252-991A-8258 (1-3414)
QY 85 SerSerLeuLysAspIleAspIleAlaArgGlu-----AsnAsnPro 98
DB 91 TCAGCATAGGAGAGAAACCCATGGCCAGACGCGAGAACCGGTCTTTCTCAACGACCG 150
QY 99 AspTyrIleLeuSerAsnLysGlnValGlyGlyValCysTyr-----ValAspLeu 115
DB 151 CAATGTACAGGAC-----CGCGTGATCTACCGGTCCACGTGAATCC 195
QY 116 Phe-----AlaGlyAspLeuLysGlyLeuLysAspIlePro 128
DB 196 TTCTACGACGCCAACACAGATGATTCGGCGATTCGCCGCGCTGATCGAGAGCTCGAC 255
QY 129 TyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGlu 148
DB 256 TACATCGCCGACCTCGCGGTGACACTCTCTGGCTGCTGCGCTCTCTAC-----CGTCCG 309
QY 149 GlyLysSerAspGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGly 168
DB 310 CCACGCGCGACGACGCTACGACATCGCCCTAGTACCGTGGCGTGCACACGACGCTACGCG 369
QY 169 ThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaVal 188
DB 370 AGCTCGCGACGCGCCCGCTTCATTCGCCAGGCCCATCGACGCGGTCTGCGGGTCAAT 429

QY 189 ValAspPheIlePheAsnHisThrSerAsnGluHisGluTyr-----AlaGlnArg 205
DB 430 ACCGAGGTGGTATCAACCAACCACTCCGACCATCCCTGGTTCATCCGCGCGCCGAC 489
QY 206 CysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAspArgMet 225
DB 490 GCGAAGAGGATCCGCGCGCCGCGAC-----TACTACGTCTGTGTCGAC-----AGC 537
QY 226 ProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe 245
DB 538 CACGAGAAATACACGAGGTACG-----CGGATCATCTTTCATCGAC----- 576
QY 246 SerGlnLeuGluAspGlyArgTyrValTyrThr----- 256
DB 577 -----ACCGAGCAGTCCCACTGGACCTGGGACCCGCTAGCCCAACAGTACTACTGGGAC 630
QY 257 ThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMet 276
DB 631 CGCTTCTATTCACACCGACCTGAACTTCGACACCCGAGGTCTCTGCGGAGGTG 690
QY 277 AlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaVal 296
DB 691 CTCGGGTGATCGCTACTGTGTCGACATGGCGCTCGACGCGCTGCGCTCGACGCGAAT 750
QY 297 AlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeu 316
DB 751 CCCTACTCTGATCGAACCGGACCGACCGACGAGGAACTTCGCGGAGACCCACCGAGTG 810
QY 317 IleArgAlaPheAsnAlaValMetArgIleAlaProAlaValPhePheLysSerGlu 336
DB 811 CTCAGGCGCATCCGCGCGGACGTGACGCGCACTATCCGACCGCATCTGCTGTCGCGGAG 870
QY 337 AlaIleValHisProAspGlnValValGlnTyrIleGly-----Gln 350
DB 871 GCCAACCATGGCGGAGAACACCGCGCGCTACTTCGCGCGGAGGATGGCGCGGAGGCG 930
QY 351 AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAla 370
DB 931 GACGATGCCACATGGCTTCCACTCCCGCTGATGCGCGCATATACATGCCATCGCC 990
QY 371 ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
DB 991 CAGGAGGATCGCTATCCGATCACCGACATCTTCGCCAGACCCCGGACATCCCGCCCAAT 1050
QY 391 ThrAlaTyrValAsnTyrValArgSerHisAspIleGlyTyrThrPheAlaAspGlu 410
DB 1051 TGCCAAATGGCGGATCTTCTCGGCAACCAACGACGAGTGAACCTGGAGATGTCACCGAC 1110
QY 411 AspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePhe 430
DB 1111 -----GACGAGCGCGACTATCTC----- 1128
QY 431 ValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThr 450
DB 1129 -----TGGAACCATATGCC 1143
QY 451 GlyAspCysArgValSer-----GlyThrAlaAlaLeuValGlyLeuAlaGln 467
DB 1144 GCCGACCGCGCGCGCTCAACCTGGGCATCCCGCGCGCTGCGCGCTGTGGAG 1203
QY 468 AspAspProHisAlaValAspArgIleLysLeuTyrSerIleAlaLeuSerThrGly 487
DB 1204 CGTGACCGCGCG-----CGCATCGAGTCTGACACGCGCTCTGTGTGTCGATCGCG 1254
QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
DB 1255 GGCACGCGCGACCTCTACTACGCGCGAGATCGGATCGGCGACCAACATCTACTACCTCGGC 1314
QY 503 AspAspAsp-----TyrSerGlnAspSerAsn----- 511
DB 1315 GACCGCGACGCGGTGCGGACCCCGATGTCAGTGTGTCGTCGCGACCGCGCGGTCTTCC 1374
QY 512 LysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeuTyr----- 529

3696	DB	TTCTACGAGGCCAACACGATGGTATCGCGACTTCGCGGGCTGATCGAGAGAGCTCGAC	3693
129	QY	TyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGlu	148
3636	DB	TACATCGCGCGACTCGGGGTGAACACTCTCTGGCTGTCTGCCCTTTCTAC-----CGGTGG	3583
149	QY	GlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGly	168
3582	DB	CCACGCCGCGACGAGCGGTACGACATCGCCCGAGTACCGTGGCGGTGCACACGCACTACGCG	3523
169	QY	ThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaVal	188
3522	DB	AGCCTCGCGACGCCCGCGGTTCATCGCGAGGCCCATCGACGCGGTCTCGGGGTGATT	3463
189	QY	ValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp-----AlaGlnArg	205
3462	DB	ACCGAGCTGGTGATCAACACACCTCCGACGAGCATCCCTGGTTCATCCGCGCCGCCAC	3403
206	QY	CysAlaAlaGlyAspProLeuPheAsnPheTyrTyrIlePheProAspArgArgMet	225
3402	DB	GCGAGAGAGGGATCGCGCCCGCGAC-----TACTACGTCTGGTTCGGAC-----AGC	3355
226	QY	ProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe	245
3354	DB	GACGAGAAATACAGGGTAGC---CGGATCATCTTCATCGAC-----	3316
246	QY	SerGlnLeuGluAspGlyArgTrpValTrpThr-----	256
3315	DB	-----ACCGAGGATGCAACTGGACCTGGACCGGTAGCCCAACAGTACTACTGGCAC	3262
257	QY	ThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMet	276
3261	DB	CGCTTCTATTCCACCAGCGCGACCTGAATTCGCACACCCGAGGTCTCTGGCGGAGGTG	3202
277	QY	AlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaVal	296
3201	DB	CTCGGGGTGATCGCTACTTGGCTGGACATGGGGGTTCGCGGCTGGCCCTTGACCGGATT	3142
297	QY	AlaPheIleTrpLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeu	316
3141	DB	CCCTACTCTATCGAACCGACGCGCACCGACGCGAGAACCTTGC CGAGAGACCCACCAAGGTG	3082
317	QY	IleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGlu	336
3081	DB	CTCAACGCGATCGCGCGAGCTGGACGCGCACTATCCCGACCGCATGTGTGGCCGAG	3022
337	QY	AlaIleValHisProAspGlnValValGlnTyrIleGly-----Gln	350
3021	DB	GCCAACAGTGGCGGAGAACACCCGGCGGTACTTCGGCGCGAGGATGGCGCGAGGCG	2962
351	QY	AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeuAla	370
2961	DB	GACCAATGCCATGCGCTTCCACTTCGCGCTGATCCCGCATGTCGCGCATGGCCATCGCC	2902
371	QY	ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis	390
2901	DB	CAGGAGGATCGCTATCCGATCACCAGCATCTTCGGCCAGACCCCGGACATCCCGGCAT	2842
391	QY	ThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPheAlaAspGlu	410
2841	DB	TGCCAATGGCGCATCTTCTGCGCAACACGACGAGCTGACCTCCGAGATGGTGCACCGAC	2782
411	QY	AspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePhe	430
2781	DB	-----GACGAGCGCGACATATCTC-----	2764
431	QY	ValAsnArgPheAspGlySerPheAlaArgIlyValProPheGlnTyrAsnProSerThr	450
2763	DB	-----TGGAACCATCATGCC-----	2749
451	QY	GlyAspCysArgValSer-----GlyThrAlaAlaAlaLeuValGlyLeuAlaGln	467
2748	DB	GCCGACCGCGCGCGCTCAACTCTGGGCGATCCGCGCGGCTTGGCGCGCTGGCGGTGGAG	2689

RESULT 3
 US-09-252-991A-8065/c
 / Sequence 8065 Application US/09252991A
 / Patent No. 6551799
 / GENERAL INFORMATION:
 / APPLICANT: Marc J. Rubenfield et al.
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 / TITLE OF INVENTION: ASSIGNORS FOR DIAGNOSTICS AND THERAPEUTICS

[illegible]

QY	468	AspAspProHisIaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly	487
Db	2688	CGTGACCGCGG-----CGATCGACTCTGCACAGCCTGCTGTTGTCGATCCCG	2638
QY	488	GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn	502
Db	2637	GGCACCGCCACCTGTACTACGGGACGAGATCGGATGGCGGCACACATCTACTCTGGC	2578
QY	503	AspAspAsp-----TrpSerGlnAspSerAsn-----	511
Db	2577	GACCGGACGGCGTGGCGACCCGCGATGCAGTGTGGTGGACGCACCGCGGCTTCTCC	2518
QY	512	LysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeuTyr-----	529
Db	2517	CGCGCCGACCGCGAAGCTGGT---CTGCCCGCATCTCTGACCCGCTGTAGCGGTAC	2461
QY	530	-----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyr	543
Db	2460	CAGACGATCAACGTGAGGCGGCGGCGGACCCGCATTCG-----CTGCTC	2413
QY	544	GlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArg	563
Db	2412	AACTGGATGCGCGCTGCTGCGCGTGGCGAGCCAGCAGAGAGCCCTTCGGCCCGCGCAGC	2353
QY	564	LeuValThrPheAsnThrAsnAsnLysHisIleGlyTyrIle-----	578
Db	2352	CTGAAGATGCTCGCGCGGAGCAACCGACGGATCTCTCGCTACCTCGCGCAATACGCGCAG	2293
QY	579	-----ArgAsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlnTyrProGlnThrVal	596
Db	2292	GGCGACGGCAGACAGCATCTCTCGGTGGCCCAACTGTGCGCGCGGCCACAGGCGGTG	2233
QY	597	ThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis-----Asp	610
Db	2232	-----GAACTGGACCTCGCCAGCCATGCGCGCAAGTGGCGGTGGAG	2191
QY	611	LeuIleGlyGlyLysThrVal-----SerLeuAsnGlnAspLeuThrLeuGln	626
Db	2190	ATGATCGGCGCATGTGTTCCGCGCGATCGCGGAGCTGACCTACTGCTGACCCCTGCCG	2131
QY	627	ProTyrGlnValMetTrpLeuGluIleAla	636
Db	2130	CCCTACGGCTTCTACTGTTCTACCTGGCC	2101

```

RESULT 4
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Alignment Scores: 4403765
Pred. No.: 1.05e-45
Length:

Score:	509.50	Matches:	160
Percent Similarity:	41.90%	Conservative:	83
Best Local Similarity:	27.59%	Mismatches:	216
Query Match:	14.95%	Indels:	121
DB:	3	Gaps:	24

US-09-843-007A-2 (1-636) x US-09-103-840A-2 (1-4403765)

QY	117	AlaGlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThr	136
	:	:::::	:::::
Db	152683	TGGCGGCATCTCGGTGGACTCATCGATTCGCTCGACTACTCGCAGTCGCTTGGCATCGAC	152742
QY	137	TyrlLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrlala	156
	:	:::::	:::::
Db	152743	TGCATCTGGTTGCCGCGTTCTACGACTCCGC-----CTGCGCAGCGCGGTTCACGAC	152796
QY	157	ValserSeryrArgaspValasnProalaLeuGlyThrileGlyAspLeuargGluval	176
	:	:::::	:::::
Db	152797	ATTGGGACCTTCTACAAAGGTCTGCCGAATTCGGCACCGTCGACGATTTGCTCGCCCTCG	152856
QY	177	IleAlaalaLeuHisGluAlaGlyIleSerAlavalValasppheIlePheAsnHistr	196
	:	:::::	:::::
Db	152857	GTGCAGCGCCGCTCACCGCGGAGGTATCGCATCATCCGACCTGGTGATGTAATCACACC	152916
QY	197	SerAsnGluHi sglutrpalagin-----ArgCysAlaalaGlyAspproLeupheasp	214
	:	:::::	:::::
Db	152917	TGGAGTTCGCACCCCCTGGTTTCAGGAGTCCCGCGGACCCAGCAGCACCGTAGCGTGAC	152976
QY	215	AsnPheTyrtyrilePheProasp-----ArgArgMetProasppGlnTyrAspargThr	232
	:	:::::	:::::
Db	152977	-----TATTACGTGTGGAGCGACACCGAGCGCGCTACACCGAC-----	153015
QY	233	LeuArgGluilePheproaspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg	252
	:	:::::	:::::
Db	153016	GCGCGATCATCTTCTGTCGAC-----ACCGAAGATCGCAAC	153051
QY	253	Trp-----ValtrpThrPheAnserPheGlnTrp	263
	:	:::::	:::::
Db	153052	TGTCATTATCGATCTCTCGCGGACAGTTCTACTTGGCACCGGATCTTCTCTCCACCAACCG	153111

[illegible]

Db	153583	ATG----	---	153588
QY	435	AspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGlyAspCysArg	454	
Db	153586	-----TAGCCGAGTACGCCAAGGATCCACGG	153612	
QY	455	ValSer-----GlyThrAlaAlaAlaLeuValGlyLeuAlaGlnAspAspProHis	471	
Db	153613	ATGAAGCGCAATGCGGAATCGCTGCGGCTTCGCGCGCTGCACAAACACCGC---	153669	
QY	472	AlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeu	491	
Db	153670	-----AACCAAGTCGAGCTTCCGCGCGCTGCTGTGCGTGCCTCGCGGTCGCGGTC	153723	
QY	492	IleTyrLeuGlyAspGluValGlyThr-----LeuAsnAspAspAsp---	505	
Db	153724	CTCTACTACGGCGCAGAGATCGGATGGCGACGTGATCTGGTTGGTGTATCGCGACCGC	153783	
QY	506	-----TtpSerGlnAspSerAsnLysSerAspAspSerArgTrpAla	519	
Db	153784	GTGCGCATCCCGATGCAGTGGACACCGACCGCAACCGCGGTTCTCCACCGCCACCGG	153843	
QY	520	HisArg-----ProArgTyrAsnGluAlaLeuTyr-----	529	
Db	153844	GGTCGCGTGTACTCCGCGCCGACGACGACCGCGGTTTACGGGTATCAGCGCGTCAACGTC	153903	
QY	530	---AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis	548	
Db	153904	GAGCGCGAACCG---GACACCTCGAGCTCG-----CTGCTCAACTCTACTCGCACC	153951	
QY	549	MetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsn	568	
Db	153952	ATGCTGGCGGTGCGTGCGCACACCCGCGTTTTCGCTCGCGCATTCAGGAATTGGCG	154011	
QY	569	ThrAsnAsnLysHisIleLeuGlyTyrIleArg-----AsnAsnAla	582	
Db	154012	GGGTCCAAACCGCTCGGTCTGGCTTCTGCTGCTCAGTGGCGCGCGATGACGGCGACCC	154071	
QY	583	LeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGln	602	
Db	154072	GTGCTCTGTGTCAACAACCTCTCGCGATTCCTCGCAGCCCATCGAATTGGCATTCGAGCAA	154131	
QY	603	AlaMetProPheLysAlaHisAspLeuIleGly-----GlyLys	615	
Db	154132	TGGACCACTACACGCGCGTGCAGTCACTACGGGACGTGGAGTTTCCACGATCGGCCAG	154191	
QY	616	ThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIle	635	
Db	154192	GTGCGCTATCTG-----CTGACGCTGCCAGCACCGGGTTCTACTGGTTTCAGTTG	154242	

```

RESULT 5
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Alignment Scores:
 Pred. No.: 1.05e-45 Length: 4411529
 Score: 509.50 Matches: 160
 Percent Similarity: 41.90% Conservative: 83
 Best Local Similarity: 27.59% Mismatches: 216
 Query Match: 14.95% Indels: 121
 DB: 3 Gaps: 24

US-09-843-007A-2 (1-636) x US-09-103-840A-1 (1-4411529)

QY	117	AlaGlyAspLeuIsglyLeuIysAspLysIleProTyrPheGlnGluLeuGlyLeuThr	136
DB	152514	TCGGCGCATCTCGTGGAGCTCATCGACTCGCTCGACTACCTGTCAGATGGCTTGGCATCGAC	152573
QY	137	TyrLeuHisLeuMetProLeuPheIysCysProGluGlyIysSerAspGlyGlyTyrAla	156
DB	152574	TGCATCTGGTTGCCCGCTTCACGACTGCCG-----CTGGCGACGGCGGTATACGAC	152627
QY	157	ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal	176
DB	152628	ATTCCGCGACTTCTACAAGGTGCTGCCCGAATTCCGACCGCTGCACGATTTCGTGCGCCCT	152687
QY	177	IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr	196
DB	152688	GTCCGACCGCGCTCACCGCGAGGTATCCGATCATCACCGACTGGTGATGAATCACACC	152747
QY	197	SerAsnGluHisGluTrpAlaGln-----ArgCysAlaAlaGlyAspProLeuPheAsp	214
DB	152748	TCGGAGTCGCACCCCTGGTTTCAGAGTCCGCGCGGACCCACGACGACCGTACGGTGAC	152807
QY	215	AsnPheTyrTyrIlePheProAsp-----ArgArgMetProAspGlnTyrAspArgThr	232
DB	152808	-----TATTACGTTGGAGCGACACACGACGCGCTTACACCGAC-----	152846
QY	233	LeuArgGluIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg	252
DB	152847	GCCCGATCATCTTCGTCGAC-----ACCGAAGATCGCAAC	152882
QY	253	Trp-----ValTrpThrPheAsnSerPheGlnTrp	263
DB	152883	TGTCATATCGATCCTGTCGCGACAGTTCCTACTGGCACCGATTCCTCTCCACCAACCG	152942
QY	264	AspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMetLeuPheLeu	283
DB	152943	GATCTGAATACACAAACCCCGCGTGCAGAGCGCATGATCGACGTCATCCGCTTTTGG	153002
QY	284	AlaAsnIleGlyValAspIleLeuArgMetAspAlaValaPheIleTyrLysGlnMet	303
DB	153003	CTCGGCTTGGGCATCGACGGGTTTCGGTTGGACCGGTGCCCTTCTCTTTGAACCGTGAG	153062
QY	304	GlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaVal	323
DB	153063	GGACCAACCTGCGAGAACCTGCGCGGAACACACACGCTTTCTCAAGCGAGTCCGCAAGTG	153122
QY	324	MetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAspGln	343
DB	153123	GTGGACGACGAATTCGCCGCGCGGTGCTGTAGTCCGAAGCCAAATAGTGGCCGGCGCAT	153182
QY	344	ValValGlnTyrIleGlyGln-----AspGluCysGlnIleGlyTyrAsn	358
DB	153183	GTGTCGATATTCGTTGATCCCAACACCGGTGGCGACGAGTGCCACATGGCCCTTTCAC	153242
QY	359	---ProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeu---	376
DB	153243	TTCCCGCTGATCGCGGCACTTTC-----ATGGCCGTGCGCGGGAGTCCCGTTTT	153293
QY	377	---LeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsn	395
DB	153294	CCGATCTCGAGATCATCGCCGACACCCACCAATCCTGCATGGCGCAATGGGGATA	153353
QY	396	TyrValArgSerHisAspIleGlyTyrThrPhe---AlaAspGluAspAlaTyr	414

Db 406 -----ACCGACGAGCTCTACACGAGCGCGGGGTGATCTTCGTGCACACCGAGCGGTGC 459
Qy 244 GlypheSer---GlnLeuGluAspGlyArgTrpValTrpThrPheAsnSerPheGln 262
Db 460 AACTGGACGTGGGACACGAGCGCGCGGAGTACTACTGCGACCGTCTTCACACACGAG 519
Qy 263 TrpAspLeuAsnTySerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPhe 282
Db 520 CCGGACCTGAACTTCGACACCGGAGGTCAGAGCGCTCTGAGCGCGATGCGGTTC 579
Qy 283 LeuAlaLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpGln 302
Db 580 TGGCTCGACATGGCGCTCGAGCGCTTCGGCTCGAGCGGTCCCTACTCTACGAGCGT 639
Qy 303 MetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAla 322
Db 640 CCGGACACAGCGGAGAGACCTCCCGGAGAGCGACGAGATCTCAAGCGGTGCGCGC 699
Qy 323 ValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAsp 342
Db 700 TTCGTGACGACACACTACCCGACCGGGTGTCTGTACGAGCGGAACAGTGGCGGACC 759
Qy 343 GlnValValGlnTrpIleGlnAspGluCysGlnIleGly----- 356
Db 760 GACGTGGTGGAGTACTTCGGGCGGAGGAGCGTGGAGCGGACCGGTCTCGGCGCCGAG 819
Qy 357 -----TyrAsnProLeuGln-MetAlaLeuLeuTrpAsnThrLeuAla 371
Db 820 AGTCACATGGCTTCACCTTCGGGTGATGCGCGCATCTTCATGCGGTGCGCGCGAG 879
Qy 371 rArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 391
Db 880 TGGC---GCTCCGATCTCGGAGATCATGAGAGACCGCGCG-ATCCCGGAGGGCTG 935
Qy 391 rAlaTrpValAsnTyrValArgSerHisAspIleGlyTrpThrPhe---AlaAsp 410
Db 936 CAGTGGGCGATCTTCCTCGCACACACGACGAGTACCCTCGAGATGGTCACGACGA 995
Qy 410 uAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 430
Db 996 CGACCGCGCTACATGG----- 1014
Qy 430 eValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer 450
Db 1015 -----GGGAGTACCCAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1050
Qy 450 rGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAsp 470
Db 1051 -----ATCGGATCG 1092
Qy 470 chiAlaValAspArgIleGlyLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu 490
Db 1093 -----ACGACACGATCGAGCTTCCACCGCGGTGCTGCTGCTGCTGCTGCTGCTGCT 1145
Qy 490 oLeuIleTyrLeuGlyValGlyValGlyThrLeuAsnAspAspAspTrpSerGlnAsp 510
Db 1146 GTCTCTGTACTACGGGACAGATCGC-----ATGGCGCACACATCTGCTGCTGCTGCTG 1202
Qy 510 rAsn-----LysSerAspAspSerArgTrpAla----- 519
Db 1203 CGACGGCGTGCCTACCGCGATCGAGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1262
Qy 520 -----HisArgProArgTyrAsnGluAlaLeuTyr----- 529
Db 1263 CAGCGCGGACAGCTGCTGCGGACGATCCAGGACCGCGGTCTACGGCTACGAGCGT 1322
Qy 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGly 546
Db 1323 CAACTGCGAGGCGCGCTGGAGAACCCCTCTCG-----CTGCTGCACTGGAC 1370
Qy 546 uArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuVal 566
Db 1371 CCGCGCGATGATCCATCCG 1430

Qy 566 rPheAsnThrAsnLeuLysHisIleIleGlyTyrIleArg----- 579
Db 1431 CTTCCGGCGGTTCGAACCCCGCGGTGCTGCTCTAGTGGCGGAGTCCGGCGGCGAGCGGG 1490
Qy 580 AsnAsnAlaLeuAlaPheGlyAsnPheSerGlyTrpGlnThrValThrAlaHis 599
Db 1491 CGACGAGTGTCTCTCTCGGTCAACACCTGCTCCCGTTCCTCGGCGCGGT----- 1542
Qy 599 sThrLeuGlnAlaMetProPheLysAlaHis-----AspLeuIleGly----- 614
Db 1543 -GAGCTCGACCTCCGGAAGTACGAGGCGCGGTACCGGTGAGCTGATCGCGCGGTGCC 1601
Qy 615 -----LysThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGlnValMet 632
Db 1602 GTTCCCGCGGTGCGGAGCTCCCGTATCTCTGACGCTCAGCGCGGCGGCTTCTACTG 1661
Qy 632 pLeuGlu 635
Db 1662 GTTCCGCGCTC 1671

RESULT 7
US-08-528-199-5
; Sequence 5, Application US/08528199
; Patent No. 5763228
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING
; TITLE OF INVENTION: MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/528,199
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,126
; FILING DATE: 14-SEP-1995
; APPLICATION NUMBER: US 08/485,126
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 156399/1994
; FILING DATE: 16-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA-6A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; NAME/KEY: CDS

LOCATION: 1..1704
US-08-528-199-5

Alignment Scores:
Pred. No.: 7,71e-51 Length: 1704
Score: 504.00 Matches: 158
Percent Similarity: 41.89% Conservative: 95
Best Local Similarity: 26.16% Mismatches: 232
Query Match: 14.79% Indels: 120
DB: 1 Gaps: 22

US-09-843-007a-2 (1-636) x US-08-528-199-5 (1-1704)

Qy	98	ProAspTrrIle-----LeuSerAsnLys	105
Db	22	CCCGAGTGGTCCGACCGCGGTCTTCTACGAGGTCCTGGTGGTCCCTTCGGGACCCC	81
Qy	106	GlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAsp	125
Db	82	AACGCGCGCGGCG-----ACGGGTGACTTCCGGCGGCTCGCGGAG	120
Qy	126	LysIleProTyrPheGlnGluLeuLysLeuThrTyrLeuHisLeuMetProLeuPheLys	145
Db	121	AAGCTCGACTACCTGCACTGCGCGTGCAGTCCCTGGTGGTCCCGCTTCTTACG	180
Qy	146	CysProGluGlyLysSerAspGlyTyrAlaValSerSerTyrArgAspValAsnPro	165
Db	181	TCGCGG-----CTCGCGACGGGGTACGAGTCCGCGACTACACCGGATCCTCCG	234
Qy	166	AlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIle	185
Db	235	GAGATCGGCACGGTGGAGTCTCCACGCCCTTCTCGACGGCGGCACGAGCGGGATC	294
Qy	186	SerAlaValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp-----Ala	203
Db	295	CGGGTATCATCGACTTCGTATGATACACACAGTACGCGGACCCCGTGTTCAGGCC	354
Qy	204	GlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAspArg	223
Db	355	TCCCGACGATCCCGACGGCGCGTACGCGAC-----TTCTAGTCTGTGTCGAC	405
Qy	224	ArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisProGly	243
Db	406	-----ACCGACGACTTACGAGACCGCGGGGTGATCTTCGTCACACCGACCGTGC	459
Qy	244	GlyPheSer---GlnLeuGluAspGlyArgTrpValTrpThrThrPheAsnSerPheGln	262
Db	460	AACGAGAGTGGGACGACGACCGCGCGGCTACTACTGCGACCGCTTCTCCACACCG	519
Qy	263	TrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPhe	282
Db	520	CCCGACCTGAATTCGACAAACCGAAGTCCAGGACGACCATGCTGGAGGCGATGGCGTTC	579
Qy	283	LeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGln	302
Db	580	TGGCTCGACATGGGCTCGACGGTCTCGGCTCGACGCGGTGCGCTTCTACGAGCGT	639
Qy	303	MetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAla	322
Db	640	CCCGCACCAACGCGGAGAACTCCCGACGACGACGAGATGCTCAAGCGGGTCCGCGC	699
Qy	323	ValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAsp	342
Db	700	TTCGTGACGACAACTACCCCGCGGTGCTGTGACGAGGGCGACACGATGGCGGAC	759
Qy	343	GlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGly-----	356
Db	760	GACGTGGTGGATCTTCGGGCCCGGAGGAGCGTACGAGCAGCGTCTCGCGGCCGAG	819
Qy	357	-----TyrAsnProLeuGln-MetAlaLeuLeuTrpAsnThrLeuAlaLath	371
Db	820	AGTCACATGGCTTCCACTTCCCGTGTGATGCGCGCATCTTCATGGCGGTGCGCGCGAG	879

Qy	371	rArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisTh	391
Db	880	TCGC---GCTTCCCGATCTCGGAGATCAGGACAGACGCGCGG-ATCCGAGAGGCTG	935
Qy	391	rAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPhe---AlaAspG	410
Db	936	CCAGTGGGATCTTCTTCGCAACACGACGAGTACCTCAGATGGTACCGGACGA	995
Qy	410	uAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPheP	430
Db	996	GGACCGCGACTACATGGG-----	1014
Qy	430	eValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerTh	450
Db	1015	-----GGCGAGTACCAAGGACCCCGCATGAGGCCAAC-----	1050
Qy	450	rdGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspPr	470
Db	1051	-----ATCGCATCCGCGCGCGCTCGCGCTGTGTGTGCGCGCTCCCC	1092
Qy	470	oHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuPr	490
Db	1093	-----ACGACCGATCGAGCTGTTCACCGCGCTCTGTGTGTGCGCGCTCCCC	1145
Qy	490	oLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspTrpSerGlnAspSe	510
Db	1146	CGTCTGTACTACGCGACGAGATCGGC---ATGGCGACCAACATCTGGCTCGGTGACCG	1202
Qy	510	rAsn-----LysSerAspAspSerArgTrpAla-----	519
Db	1203	CGACGGCTGCTGACGCGGATGACGCGGACCCCGACCGCAACGTCGCGTTCGCGCGCG	1262
Qy	520	-----HisArgProArgTyrAsnGluAlaLeuTyr-----	529
Db	1263	CACGCGCGCAAGTCGACCTGCGAGATCCAGGACCGGCTCTACGGCTACGAGCGGT	1322
Qy	530	-----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLe	546
Db	1323	CAACGTCGAGCGCGAGCTGGAGAACCCCTCTCG-----CTGTGCACTGGAC	1370
Qy	546	uArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValTh	566
Db	1371	CGCGCGATGATCCACATCCGCGCGACGCGCGCTTCGGGTGCGGACCTTCGAGGA	1430
Qy	566	rPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg-----	579
Db	1431	CCTCGCGCGCTCGAACCCGCGGTGCTGTCTCTACTGCTGCGGAGTCCGCGGCGACG	1490
Qy	580	-AsnAsnAlaLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHi	599
Db	1491	CGACGAGTGTCTCTGCGTCAACACCTGTCCTCCGCTTCCCGACGCGGTC-----	1542
Qy	599	sThrLeuGlnAlaMetProPheLysAlaHis-----AspLeuIleGlyGly-----	614
Db	1543	-GAGCTCGACCTCGGAAGTACGAGGCGCGGTACCGGTGAGCTGATCGCGCGGTGCC	1601
Qy	615	-----LysThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGlnValMetTr	632
Db	1602	GTTCCCGCGCTCGGAGGCTCCGATATCTCTGAGCTCAGCGGCGACGCGTCTTACTG	1661
Qy	632	pLeuGluIle	635
Db	1662	GTTCCGCGCTC	1671

RESULT 8

US-08-537-002A-4
Sequence 4, Application US/08537002A
Patent No. 5773282
GENERAL INFORMATION:
APPLICANT: TSUSAKI, Keiji
APPLICANT: KUBOTA, Michio
APPLICANT: SUGIMOTO, Yoshiyuki
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR

Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnAspLysHisIleIleGlyTyrIleArgAsn 580
Db 1369 CGGGAGCCCTACCTCTCTCCCGTGAGAACCGCGCGCTCTCGCTACCTAGGGAG 1428
Qy 581 Asn-----AlaLeuAlaPheGlyAsnPheSerGlyTyrProGlnThrValThr 597
Db 1429 CACGAGGGGAGCGGCTCTGTGTGGCCAACTCTCCCGCTACACCGAGCCCTTT--- 1485
Qy 598 AlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyThrVal 617
Db 1485 ----- 1485
Qy 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
Db 1486 -----GACCTCCCTTGAGGCGCTTACCAA 1509

RESULT 9

US-08-863-010-4
Sequence 4, Application US/08863010
Patent No. 6087146
GENERAL INFORMATION:
APPLICANT: TSUSAKI, Keiji
APPLICANT: KUBOTA, Michio
APPLICANT: SUGIMOTO, Toshiyuki
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/863,010
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/537,002
FILING DATE: 29-SEP-1995
APPLICATION NUMBER: JP 260984/1994
FILING DATE: 01-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP No. 6087146 yet received
FILING DATE: 08-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TSUSAKI=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-863-010-4

Alignment Scores:
Pred. No.: 1,66e-48 Length: 2889
Score: 488.00 Matches: 147
Percent Similarity: 42.39% Conservative: 87

Best Local Similarity: 26.63% Mismatches: 202
Query Match: 14.32% Indels: 116
DB: 3 Gaps: 22
US-09-843-007A-2 (1-636) x US-08-863-010-4 (1-2889)
Qy 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnLeuGlyLeuThrTyr 137
Db 82 GGGACTTTGAGGCGCTGAGCGGAGCTTCCCTACCTGAGGAGCTCGGGGTCAACACC 141
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157
Db 142 CTCTGGCTCATGGCCTTCTTCCAGTCCCTCC-----TTGAGGACGACGGGTACGATATC 195
Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 196 TCCGACTACTACAGATCTCCCGCTCCAGCGACCTCGGAGACTTC-----ACCGTG 249
Qy 178 AlaAlaLeuHisGluAlaGlyTleSerAlaValAspPheIlePheAsnHisThrSer 197
Db 250 GACGAGCCCGACCGCGGGGATGAGGTGATCATTTAGCTCGTCTCGTGAACACCACTCC 309
Qy 198 AsnGluHisGluTTPAlaGlnArgCys---AlaAlaGlyAspProLeuPheAspAsnPhe 216
Db 310 ATTGACCACCTTGTTCAGGAGCGGAGGAGCGGATAGCCCATGCGGAC----- 363
Qy 217 TyrTyrIlePheProAspArgA-gMetProAspGlnTyrAspArgThrLeuArgGluIle 236
Db 364 TGGTACGTGTGAGCGGAC-----ACCCCGGAGAGTAC---AAGGGGGTCCGGGTCTATC 414
Qy 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTTPVal----- 254
Db 415 TTCAAGGAC-----TTTGAAACCTCCCACTGACCTTTGAC 450
Qy 255 -----TTPThrThrPheAsnSerPheGlnTTPAspLeuAsnTyr 267
Db 451 CCGGTGGCCAAAGGCTACTACTGGCACCGCTTCTACTGGCACGACCGGACCTCACTGG 510
Qy 268 SerAsnProTTPValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287
Db 511 GACAGCCCGAGGTGAGAGGCGCATCCACGAGTCATGTCTTCTGGGCGGACCTGGGG 570
Qy 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTTPLysGlnMetGlyThrSerCys 307
Db 571 GTGACGCGCTTCCGCTGGAGCCATCCCTTACCTTACGAGCGGAGGAGGACCTCTGTC 630
Qy 308 GluAsnLeuProGlnAla---HisAlaLeuIleArgAlaPheAsnAlaValMetArgIle 326
Db 631 GAGAACCTCCCGGAGACCATTTGAGCGGTGAGCGCTGAGAGGCGCTCGGAGGAGCGC 690
Qy 327 AlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGln 346
Db 691 TACGCGCCCGGAGAGATCTCTCTCGCGGAGGCCAACATGTGGCCGAGGACCTCTCCC 750
Qy 347 TyrIleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeu 365
Db 751 TACTTCGGGAGCGGAGCGGGTCCACATGGCCCTACCACTTCCCTCTGATGCCCGGATC 810
Qy 366 TrpAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHis 385
Db 811 TTCATGGCCCTTAAGCGGGAGGACCGGGGTCTCCCATTTGAACCATGCTCAAGAGGCGGAG 870
Qy 386 AsnLeuProGluHisThrAlaTTPValAsnTyrValArgSerHisAspAspIleGlyTrp 405
Db 871 GGGATCCCGGAAACCGCGGAGTGGGCGCTCTCTCCGCGCAACACGACGAGCTACCGCTG 930
Qy 406 ThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425
Db 931 GAGAGGCTCACGGAG-----GAGGAGCGGGAGTTC 960
Qy 426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445
Db 961 ATGTACGAG-----GCTACGCGCCCGGACCCCAAGTTCCGC 996

446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeu 465
997 ATCAAC-----CTGGGATCCGCGCGCTCATGCCCTC 1032
466 AlaGlnAspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeuSer 485
1033 CTCGGGGCGACCGCAGG-----CGGTACAGCTCTCTACCGCCCTCTCTCTAC 1083
486 ThrGlyGlyLeuProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 504
1084 CTAAAGGCGACGCCCATCTGCTACTACCGGGAGAGATCGGC---ATGGGGGCAACAC 1140
505-----AspTrpSerGlnAspSerAsnLys 512
1141 TTCCTCGGGACCGGACCGGTGTCAGACCCCATCGAGTGGTCCCAAGACCGCATCTC 1200
513 SerAspAspSerArgTrpAlaHisArg-----ProArgTyrAsnGluAlaLeuTyr 529
1201 GCCTTCTCCCGCGCCCTACACCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
530-----AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541
1261 AGTACCACTTCGTCTACCGTGGGCGCCGCGGGAACCCCTCTCTCTCTCTCTCT 1308
542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg---PheAsp 560
1309 CTCCTGAGCTTCAACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1368
561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
1369 CGGGGAGCTTCACT 1428
581 Asn-----AlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThr 597
1429 CACGAGGGGAGCGGCT 1485
598 AlaHisThrLeuGlnAlaMetProPheLeuAlaHisAspLeuIleGlyLysThrVal 617
1485-----GACCTCCCTTGGAGGCTTACCA 1509
618 SerLeuAsnGlnAspIleThrLeuGlnProTyrGln 629
1486-----GACCTCCCTTGGAGGCTTACCA 1509

RESULT 10
US-09-024-429-4
Sequence 4, Application US/09024429
Patent No. 6165768
GENERAL INFORMATION:
APPLICANT: TSUSAKI, Keiji
APPLICANT: KUBOTA, Michio
APPLICANT: SUGIMOTO, Toshiyuki
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
OPERATING SYSTEM: PC-DOS/MS-DOS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/024,429
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/537,002

FILING DATE: 29-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 260984/1994
FILING DATE: 01-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 255829/1995
FILING DATE: 08-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: TSUSAKI=18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-024-429-4
Alignment Scores:
Pred. No.: 1,668-48 Length: 2889
Score: 488.00 Matches: 147
Percent Similarity: 42.39% Conservative: 87
Best Local Similarity: 26.63% Mismatches: 202
Query Match: 14.32% Indels: 116
DB: 3 Gaps: 22
US-09-843-007A-2 (1-636) x US-09-024-429-4 (1-2889)
Qy 118 GlyAspLeuGlyLeuLysAspLysLeuProTyrPheGlnGluLeuGlyLeuThrTyr 137
Db 82 GGGGATTTGGAGGCGCTGAGCGGAGCTTCCCTACCTGGAGGAGCTCGGGGTCAACACC 141
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
Db 142 CTCGGCTCATGCCCTTCTTCCAGTCCCC-----TTGAGGACGACGCGGTACGATATC 195
Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 196 TCGACTATACAGATCTCTCCCTCCAGGAGGAGCCCTGGAGGACTTC-----ACGTG 249
Qy 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
Db 250 GACGAGGCGCCACGCGCGGAGTGAAGGTGATCATTTGAGCTCGTCTGAAACCCACCTCC 309
Qy 198 AsnGluHisGluTyrAlaGlnArgCys---AlaAlaGlyAspProLeuPheAspAsnPhe 216
Db 310 ATTGACCACTTGGTTCCAGGAGGAGCGGAGCGGAGTACGCCCATCGGGAC----- 363
Qy 217 TyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236
Db 364 TGGTACGTGTGAGGCGAC-----ACCCCGGAGAGTAC---AAGGGGGTCCGGGTGATC 414
Qy 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpVal----- 254
Db 415 TTCAAGGAC-----TTTGAACCTCCCACTGGACCTTTGAC 450
Qy 255-----TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyr 267
Db 451 CCGGTGGCCAGGCGCTACTCTGGACCGCTTCTACTGGACCGCTTCTACTGGACCGCTTCT 510
Qy 268 SerAsnProTyrValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287
Db 511 GACAGCGCCCGAGGTGGAGAGGCCATCCACCGAGGTCTATGTTCTTCTGGCGGACCTGGG 570
Qy 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTyrLysGlnMetGlyThrSerCys 307
Db 571 GTGGAGGCTTCCGACTGAGACGCCATCTCTACCTACCTCTACGAGCGGAGGAGGACCTCT 630

308 GluAsnLeuProGlnAla---HisAlaLeuIleArgAlaPheAsnAlaValMetArgIle 326
631 GAGAACTCCCGAGACCACTTGGGGGAGGAGCGCTGAGGAGGAGCGGC 690
327 AlaAlaProAlaValPhePheIleValHisProAspGlnValValGln 346
691 TACGGCCCGGGAAGATCTCTCGCGGAGGAGCAATGTGGCGGAGGAGACCTCC 750
347 TyIleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeu 365
751 TACTTCGGGAGCGGAGCGGGTCCACATGGCTCACTTCCCTGATGCCCCGGATC 810
366 TrpAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHis 385
811 TTCTATGGCCCTAAGCGGGAGGAGCGGGGTCCCAATGAACCACTCTCAAGGAGCGGAG 870
386 AsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspIleGlyTrp 405
871 GGGATCCCGGAAACCGCCAGTGGCCCTTCTCCGCAACCAACGAGGAGTCACTG 930
406 ThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425
931 GAGAAGGTCAAGGAG---GAGGAGCGGAGTTC 960
426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445
961 ATGTAGGAG-----GCTAGCCCGCCGACCCCAAGTTCGCG 996
446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeu 465
997 ATCAAC-----CTGGGAGTCCCGCGCGCTCATGCCCTC 1032
466 AlaGlnAspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeuSer 485
1033 CTCGGGGGAGCGGAGG-----CGTACGAGCTCTCACCGCCCTCTCTCTCACC 1083
486 ThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAsp--- 504
1084 CTAAGGGGACGCCCATCTCTACTACGGGACGAGATCGGC---ATGGGGGACACCCC 1140
505 -----AspTrpSerGlnAspSerAsnIlys 512
1141 TTCCTCGGGGACCGGAGCGGTGTCAGGACCCCGCATGTCAGTGTCCCAAGCGCATCGTC 1200
513 SerAspAspSerArgTrpAlaHisArg-----ProArgTyrAsnGluAlaLeuTyr 529
1201 GCCTTCTCCCGGCGCCCTTACCGCCCTCTTCTTCTCCCGCGTGTGAGCGGGGCGCTTAC 1260
530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541
1261 AGCTACCACTTCGTCAACGTGAGGAGCGGCGGAGAAACCCCATCC----- 1308
542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg---PheAsp 560
1309 CTCCTGAGCTTCAACCGCGGCTTCTCGCCCTGAGGAGAACCGACGACGACGATCTTCGGC 1368
561 GlyGlyArgLeuValThrPheAsnThrAsnAsnIlysHisIleIleGlyTyrIleArgAsn 580
1369 CGGGGAGCCTCACCTTCTCCCGTGGAGAACCGCGCGCTCTCGCTACTCTGAGGAG 1428
581 Asn-----AlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThr 597
1429 CACGAGGGGAGCGGGTCTGTGTGGCCCACTCTCTCCGCTACACCCAGGCGCTTT--- 1485
598 AlaHisThrLeuGlnAlaMetProPheIysAlaHisAspLeuIleGlyGlyThrVal 617
1485 ----- 1485
618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
1486 -----GACCTCCCTTGGAGGCGCTTACCAA 1509

RESULT 11
US-08-537-002A-5
; Sequence 5, Application US/08537002A
; Patent No. 5773282
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,002A
; FILING DATE: 29-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP No. 5773282 yet received
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TSUSAKI=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-528-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: DOUBLE
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Thermus aquaticus
; INDIVIDUAL ISOLATE: ATCC 33923
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..540
; IDENTIFICATION METHOD: E
; NAME/KEY: mat peptide
; LOCATION: 541..3429
; IDENTIFICATION METHOD: S
; NAME/KEY: 3'UTR
; LOCATION: 3430..3600
; IDENTIFICATION METHOD: E
US-08-537-002A-5
Alignment Scores:
Pred. No.: 2,41e-48 Length: 3600
Score: 488.00 Matches: 147
Percent Similarity: 42.39% Conservative: 87
Best Local Similarity: 26.63% Mismatches: 202
Query Match: 14.32% Indels: 116
DB: 1 Gaps: 22
US-09-843-007A-2 (1-636) x US-08-537-002A-5 (1-3600)

QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
DB 622 GGGGACTTTGAGGGGCTTGAGGGGGAAGCTTCCCTACTGGAGAGCTCGGGGTCAACACC 681
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157
DB 682 CTCGTGCTCATGCTCTTCTCCAGTCCGCC-----FTGAGGAGACGAGCGGTACGATATC 735
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValle 177
DB 736 TCCGACTACTACGAGATCTCCCGCTCCACGGAGCCCTCGAGGACTTC-----ACCGTG 789
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197
DB 790 GACGAGCCCAAGCGCGGGGATGAAGGTGATCATGTGAGCTCGTCTGAACACACACTTC 849
QY 198 AsnGluHisGluTrpAlaGlnArgCys-----AlaAlaGlyAspProLeuPheAspPhe 216
DB 850 ATGACACACCTTGGTTCAGGAGCGGAGAGCCGATAGCCCATGCGGAGC-----903
QY 217 TyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236
DB 904 TGGTAGCTGTGAGCGAC-----ACCCGAGAGATAC---AAGGGGTCCGGGTCTATC 954
QY 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpVal-----254
DB 955 TTCAGGAC-----TTTGAACCTCCCAACTGGACCTTTGAC 990
QY 255 -----TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyr 267
DB 991 CCGGTGGCCAAAGCCTACTACTGGCACCGCTTCTACTGGCACAGCCCGACCTCAACTGG 1050
QY 268 SerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287
DB 1051 GACAGCCCGAGGTGGAGAGGCCATCCACAGGTCTGCTTCTTGGGCCGACCTGGGG 1110
QY 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCys 307
DB 1111 GTGAGCGGTCTCGCTCGAGCGCTCCCTCTACTCTAGCGGGGAGGAGCTCTCTGC 1170
QY 308 GluAsnLeuProGlnAla---HisAlaLeuIleArgAlaPheAsnAlaValMetArgIle 326
DB 1171 GAGAACCTCCCGAGACCATTTGAGCGGTGAGAGCGCTGAGGAGGCCCTGGAGGAGCGC 1230
QY 327 AlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAspGlnValGln 346
DB 1231 TAGCGCCCGGGAAGATCTCTCTCGCGGAGCCCAACATGTGCGCGGAGGAGACCTCCGCC 1290
QY 347 TyrIleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeu 365
DB 1291 TACTTCGGGAGCGGAGCGGGTCCACATGGCTTACAACTTCCCTCGATGCCCGGATC 1350
QY 366 TrpAsnThrIleAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHis 385
DB 1351 TTCATGGCCCTAAGCGGGGAGGACCGGGGTCCCATTTGAACCATGCTCAAGGAGCGGAG 1410
QY 386 AsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrp 405
DB 1411 GGGATCCCGAARACCCCGAGTGGCCCTTCTCTCCGCAACACGACGAGTCAACCTG 1470
QY 406 ThrPheAlaAspGluAspAlaValTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425
DB 1471 GAGAGGTTCACGGAG-----GAGGAGCGGAGGTTC 1500
QY 426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445
DB 1501 ATGTACGAG-----GCCATGCCCGCCGACCCCAAGTTCGC 1536
QY 446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeu 465
DB 1537 ATCAAC-----CTGGGATCGCGCGGCTCATGCCCTC 1572
QY 466 AlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSer 485

DB 1573 CTCGGGGGACCGCAGG-----CGGTACGAGTCTCTACCGCCCTCTCTCTCACC 1623
QY 486 ThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAsp---504
DB 1624 CTAAAGGGGACCGCCATCGTCTACTACGGGGACGAGATCGGC-----ATGGGGGACAAACCC 1680
QY 505 -----AspTyrSerGlnAspSerAsnLys 512
DB 1681 TTCCTCGGGGACCGAAGCGGTGTGAGACCCCAATCGAGTGTCCCAAGACCGCATCGTC 1740
QY 513 SerAspAspSerArgTrpAlaHisArg-----ProArgTyrAsnGluAlaLeuTyr 529
DB 1741 GCCTTCTCCCGCGCCCTTCT 1800
QY 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541
DB 1801 AGTACACCTTCGTCAACGTGAGCGCCGAGCGGGAACCCCACTCC-----1848
QY 542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg---PheAsp 560
DB 1849 CTCCTGAGCTTCAACCGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1908
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
DB 1909 CGGGGGAGCTTCAACCTTCT 1968
QY 581 Asn-----AlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThr 597
DB 1969 CACGAGGGGAGCGCGTCTCTGTGGTGGCCAACTCTCTCTCTCTCTCTCTCTCTCTCT 2025
QY 598 AlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyThrVal 617
DB 2025 -----2025
QY 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
DB 2026 -----GACCTCCCTTGGAGCGCTTACCA 2049

RESULT 12
US-08-863-010-5
; Sequence 5, Application US/08863010
; Patent No. 6087146
; GENERAL INFORMATION:
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: KOBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,010
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,002
; FILING DATE: 29-SEP-1995
; APPLICATION NUMBER: JP 260984/1994
; FILING DATE: 01-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP No. 6087146 yet received

QY 426 LeuAsnArgPheValAsnArgPheAspGlySerPheAlaValGlyValPropheGln 445
Db 1501 ATCTAGAG-----GCTACGCCCGCCAGCCCAAGTTCGCG 1536
QY 446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaValGlyLeu 465
Db 1537 ATCAAC-----CTGGGATCCCGCGCGCTCATGCCCCCTC 1572
QY 466 AlaGlnAspProHisAlaValAspArgGlyLeuLeuTyrSerIleAlaLeuSer 485
Db 1573 CTGGGGGGGACCGCAGG-----CGTACGAGCTCTCTACCGCCCTCTCTCTCACC 1623
QY 486 ThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAsp 504
Db 1624 CTAAGGGGACCGCCCATCGTCTACTACGGGGGAGATCGGC---ATGGGGGACACCC 1680
QY 505 -----AspTyrSerGlnAspSerAsnLys 512
Db 1681 TTCTCTGGGACCGGAACGGTGTTCAGGACCCCATCGAGTGTCTCCAGACCGCATCTC 1740
QY 513 SerAspAspSerArgTrpAlaHisArg-----ProArgTyrAsnGluAlaLeuTyr 529
Db 1741 GCCTTCTCCCGCCCGCCCTTACCACCGCTCTCTCTCCCGCGTGTGAGCGGGGCTTAC 1800
QY 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541
Db 1801 AGTACACCTCTCTCAACGTGGAGGCGCCCGGAGAAACCCCATCTCC----- 1848
QY 542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg-----PheAsp 560
Db 1849 CTCTGAGCTTCAACCGCGCTCTCTCGCCCTCAGGAACCGACGACCGCAAGATCTTCGGC 1908
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
Db 1909 CGGGGGAGCGCTACCTCTCTCCCGGGAGAACCGCGCGTCTCTCGCTACCTGAGGAG 1968
QY 581 Asn-----AlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThr 597
Db 1969 CAGAGGGGAGCGGCTCTGTGGTGGCGCACTCTCCGCTACACCGAGCGCTTT--- 2025
QY 598 AlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyLysThrVal 617
Db 2025 -----GACCTCCCTTGGAGGCTACCA 2049
QY 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
Db 2026 -----GACCTCCCTTGGAGGCTACCA 2049
RESULT 14
US-09-107-532A-2079
; Sequence 2079, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A. Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO. 2079:
SEQUENCE CHARACTERISTICS:
LENGTH: 1728 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1728
SEQUENCE DESCRIPTION: SEQ ID NO: 2079:
US-09-107-532A-2079
Alignment Scores:
Pred. No.: 9,6e-39 Length: 1728
Score: 404.50 Matches: 156
Percent Similarity: 40.97% Conservative: 105
Best Local Similarity: 24.43% Mismatches: 277
Query Match: 11.87% Indels: 93
DB: 28
US-09-843-007A-2 (1-636) x US-09-107-532A-2079 (1-1728)
QY 34 SerGluAspTyrArgGlnPheSerArgMetAspThrHisPheProLysLeuMetAsn 53
Db 1 TCATACTCTGGAGACGCTCCCAAAAAGAAAGGGGAAT-----CCGATGTATACAAA 54
QY 54 GluLeuAspSerValTyrGlyAsnAsnGluAlaLeuProMetLeuGluMetLeuLeu 73
Db 55 CTTTTCACAAAGCATTTATCGGACTCAGAT---TTGTCTGAGAAAACCGCTCTATGATT 111
QY 74 AlaGlnAlaTrpGlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAla 93
Db 112 GAACAAAAGTCAGTGTGTTTATCAGCAGAGAAAGGCCGCGGCAACTG----- 162
QY 94 ArgGluAsnAsnProAspTyrIleLeuSerAsnLysGlnValGlyValCysTyrVal 113
Db 163 -----CATTTG-----GACGAGAAGATATTTTCTTGATTACTTATGGC 201
QY 114 AspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIlePro----- 128
Db 202 GATCAATTT-----TTTGAAGAGCAGCAACCACTTTGCGTACATTT 243
QY 129 -----TyrPheGlnGluLeuGlyLeuThrTyr-----LeuHisLeuMetProLeu 143
Db 244 AGAAATTTTATCAAAAGTATTTGGCAGATACATTTAGATTGTGATCTTCTTGCAATTT 303
QY 144 PheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyrArgAspVal 163
Db 304 TTCCCATAT-----ACATCTGATGATGTTTCTCCGTTGTAATTTATCAGGCAATC 354
QY 164 AsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAla 183
Db 355 AACGACCGTTTGGGCACATGGGAAGATATAGAAATG-----AGAAGC 399
QY 184 GlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHisGluTrpAla 203
Db 400 GAACACAGGTTAATGTTGCTGACTTTTGTCTCAATCATATGCTGCGGAAAGTACTTGTT 459


```
Db 403 TATATTGGCCAGATCTCTAAA---CCGGAT----- 429
QY 238 ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGly----- 251
Db 430 ---GGATCTGAACCTTAATACTGGGAAGATCTTTTAATGGTTCAACTGGGAGTTTGAC 486
QY 252 -----ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyr 267
Db 487 GAATCGCACTAAGCAATACTATTTCCATTTATTAGCAAAAGCAGCCAGATTATTAATGG 546
QY 268 SerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287
Db 547 GAAATCTCCAGATGTAAGCAACAGCTGTTTGAATGATGAATGGTGGTTGAAAAGGT 606
QY 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCys 307
Db 607 ATTGACGGATTTAGAGTTGATGTCATTTACTATTAATAAGAAATTTTGAAGCAGGAGAT 666
QY 308 GluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetAlaGlyIleAla 327
Db 667 TTACCTGTACTGTGTCGCAAAAATTTGCTCCAGCATTTTGATGTAGATATGAAT----- 720
QY 328 AlaProAlaVal-----PhePheLysSerGluAlaIle----- 338
Db 721 CAGCCAGGAATACAGAATGCTCCAGAAATGAAGATAAATCGTTAAGTCGGTATGAC 780
QY 339 -----ValHisProAspGlnValValGlnTyrIleGly 349
Db 781 ATTATGACTGTAGCGAGGCTAATGCTTACTCTTAATGATGCTGAAGATGGGTAGGA 840
QY 350 GlnAspGluCysGlnIleGlyTyrAsnProLeu-----GlnMetAlaLeuLeuTyr 365
Db 841 GAAGAAAATGGGAAA-----TTTAATATGATATTCCAGTTTGAACATCTTGGTTATGG 894
QY 367 AsnThrLeuAlaThrArg---GlnValAsnLeuLeuHisGlnAlaLeuThrTyrArgHis 385
Db 895 AGTACTGGCGATACGAAATTCGATGTTAAATCCCTATAAACAAAGCTTAAATCGTTGGCAA 954
QY 386 AsnLeuProGluHisThrAlaTrp---ValAsnTyrValArgSerHisAspAsp----- 402
Db 955 AAGCAACTAGAAATATGAGTTGGATGCTTTATTATTCGAAACCATTGATCAACACCGT 1014
QY 403 -----IleGlyTrpThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr 420
Db 1015 CGTGTTCACACTGG-----GGTGATGATAAAATTAATTTGATGATCAATCAGCAACT 1065
QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
Db 1066 AGTCAC-----GCTACTGCCTACTTTTACAA-----CAG 1095
QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
Db 1096 GGCACACCTTTTATTACCAAGTCAAGAA----- 1125
QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAsp-----Arg 475
Db 1126 -----ATAGGTATGACATAAATATTCATTGTAAGCAATTGAAAGTTTCAACGATGTCGA 1179
QY 476 IleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGly 495
Db 1180 GTGAAACTGATATCAATAGTCAAAAAGAGGTGGA---GATGTCAATCAATTAATACTA 1236
QY 496 AspGluValGlyThrLeuAsnAspAspAsp-----TyrSerGlnAsp 509
Db 1237 GATAAATATAAATAAGAAAACCGAGACCAATGCAAGACTCCAATCCAATGGAATTAATCT 1296
QY 510 SerAsnLys-----SerAspAspSerArgTyrAlaHis---ArgProArgTyrAsnGlu 526
Db 1297 ATCAATGCTGGATTCTACTGCTAGGCAATGTTTCATGTAAACCTTAACCTATACAGAA 1356
QY 527 AlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeu 546
```

```
Db 1357 ATTAATGTTAAACACAACTAAATGATTAAGTTTTCGATACTTTCTTATTATAAAGCGTTA 1416
QY 547 ArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThr 566
Db 1417 -----ATTCAACTAAAAAACTGATTTGATTTGATTTACACCTACGTTAAGTTTAATATG 1467
QY 567 PheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg-----AsnAsnAlaLeu 583
Db 1468 GTCGATGCTGAAAATAAAGCAGGTTTTTTCATATACACGCACATTTAAAAACAATACTGTA 1527
QY 584 LeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAla 603
Db 1528 TTAATTTGAGCCCAATCTCAAAATGAAGTATCAGAACTAAAC----- 1569
QY 604 MetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn----- 620
Db 1570 CTACCTTTTGAATTAGAT-----ATTTCATCTGTAGATATAAAATTCATATTATCAC 1623
QY 621 ---GlnAspLeuThrLeu-----GlnProTyrGlnValMetTrpLeuGluIle 635
Db 1624 TTAATATGATATAAAATTTAGACCATATTAACCTTATGAATCATTCGTCGTTGAATA 1680
```

Search completed: November 8, 2003, 23:30:29
Job time : 2977 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2003, 21:58:19 ; Search time 486 Seconds
(without alignments)
4175.821 Million cell updates/sec

Title: US-09-843-007A-2
Perfect score: 3408
Sequence: 1 MLPTQGVGLILYKTRIL.....VSLNQDLTLPYQVMLEIA 636

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2141354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xl
-Q=/cgn2/1/USPTO_Spool/p/US09843007/runat_07112003_170422_25467/app_query.fasta_1.775
-DB=Published Applications NA -QFWT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09843007@cgn2.1.107@runat_07112003_170422_25467
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOP -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubnpa/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubnpa/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubnpa/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubnpa/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubnpa/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubnpa/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubnpa/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubnpa/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubnpa/US09E_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubnpa/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubnpa/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubnpa/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubnpa/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubnpa/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	3408	100.0	2914	9 US-09-843-007-1 Sequence 1, Appli

2	541.5	15.9	3303	9 US-09-815-242-7764 Sequence 7764, Ap
3	529.5	15.5	1716	14 US-10-156-761-2790 Sequence 2790, Ap
4	529.5	15.5	1794	10 US-09-738-626-2527 Sequence 2527, Ap
5	529.5	15.5	3309400	10 US-09-738-626-1 Sequence 1, Appli
6	529.5	15.5	9025608	14 US-10-156-761-1 Sequence 1, Appli
7	511.5	15.0	1749	14 US-10-156-761-7374 Sequence 7374, Ap
8	338	9.9	16592	8 US-08-781-986A-53 Sequence 53, Appli
9	331	9.7	1620	14 US-10-081-872-153 Sequence 153, App
10	328.5	9.6	1650	9 US-09-815-242-8494 Sequence 8494, Ap
11	323.5	9.5	1638	9 US-09-815-242-4455 Sequence 4455, Ap
12	313.5	9.2	1782	14 US-10-061-269-13 Sequence 13, Appli
13	300.5	8.8	1596	14 US-10-081-872-191 Sequence 191, Appli
14	300	8.8	1431	14 US-10-081-872-171 Sequence 171, App
15	300	8.8	1596	14 US-10-081-872-167 Sequence 167, App
16	299.5	8.8	1587	14 US-10-081-872-159 Sequence 159, App
17	299	8.8	1596	14 US-10-081-872-173 Sequence 173, App
18	298	8.7	1683	10 US-09-974-300-655 Sequence 655, App
19	295.5	8.7	1596	14 US-10-081-872-157 Sequence 157, App
20	295.5	8.7	1599	14 US-10-156-761-1713 Sequence 1713, Ap
21	294.5	8.7	9025608	14 US-10-156-761-1 Sequence 1, Appli
22	294.5	8.6	1572	14 US-10-081-872-185 Sequence 185, App
23	292.5	8.6	1596	14 US-10-081-872-189 Sequence 189, App
24	292	8.6	1623	14 US-10-081-872-161 Sequence 161, App
25	291.5	8.6	1596	14 US-10-081-872-183 Sequence 183, App
26	287	8.4	1689	10 US-09-974-300-657 Sequence 657, App
27	283.5	8.3	1794	14 US-10-061-269-11 Sequence 11, Appli
28	283	8.3	1848	14 US-10-156-761-5957 Sequence 5957, Ap
29	280	8.2	1704	14 US-10-061-269-15 Sequence 15, Appli
30	276.5	8.1	1656	9 US-09-815-242-6367 Sequence 6367, Ap
31	274.5	8.1	1524	14 US-10-081-872-95 Sequence 95, Appli
32	274.5	8.1	1803	12 US-10-223-277-4 Sequence 4, Appli
33	274.5	8.1	1803	12 US-10-223-277-5 Sequence 5, Appli
34	274.5	8.1	2931	14 US-10-081-872-97 Sequence 97, Appli
35	273	8.0	4888	10 US-09-070-927A-307 Sequence 307, App
36	271.5	8.0	1536	14 US-10-081-872-99 Sequence 99, Appli
37	271.5	8.0	1703	10 US-09-974-300-644 Sequence 644, App
38	265.5	7.8	1659	12 US-10-223-277-7 Sequence 7, Appli
39	265.5	7.8	1659	12 US-10-223-277-8 Sequence 8, Appli
40	263.5	7.7	1653	9 US-09-815-242-9889 Sequence 9889, Ap
41	262.5	7.7	3354	12 US-09-814-353-21784 Sequence 21784, A
42	262.5	7.7	30246	8 US-08-781-986A-56 Sequence 56, Appli
43	260	7.6	1656	12 US-10-223-277-1 Sequence 1, Appli
44	260	7.6	1656	12 US-10-223-277-2 Sequence 2, Appli
45	253.5	7.4	2863	12 US-10-050-763-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-843-007-1

; Sequence 1, Application US/09843007

; Patent No. US20020092040A1

; GENERAL INFORMATION:

APPLICANT: Kossmann, Jens

Buttcher, Volker

Weish, Thomas

TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE

OF FACILITATING THE SYNTHESIS OF LINEAR

ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND

MICROORGANISMS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/843,007
 FILING DATE: 26-Apr-2001
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/737,752
 FILING DATE: <Unknown>
 APPLICATION NUMBER: DE P 44 47 388.5
 FILING DATE: 22-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: James P. Haley, Jr., Esq.
 REGISTRATION NUMBER: 27,794
 REFERENCE/DOCKET NUMBER: GFB-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 596-9000
 TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 2914 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Neisseria polysacchara

IMMEDIATE SOURCE:

LIBRARY: genomic library in pBluescriptII SK

CLONE: pNB2

FEATURE:

NAME/KEY: CDS
 LOCATION: 957..2867
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-843-007-1

Alignment Scores:

Pred. No.:	Score:	Length:
0	3408.00	2914
100.00%	Matches:	636
100.00%	Conservative:	0
100.00%	Mismatches:	0
100.00%	Indels:	0
9	Gaps:	0

US-09-843-007A-2 (1-636) x US-09-843-007-1 (1-2914)

```

QY      1  MetLeuThrProThrGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20
DB      957  ATGTTGACCCCGCAGCAGCTCGTGGTTGATTTTACAGTACTCAAAACACGCATCTTG 1016

QY      21  AspIleTyrThrProGlnArgAlaGlyIleGlyLysSerGluAspTrpArgGlnPhe 40
DB      1017  GACATCTACACGCCCGACAGCGCGCGCATCGAAAATCCGAGAGCTGGCGGCAGTTT 1076

QY      41  SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
DB      1077  TCGCGCGCATGATACGATTCCTCCCAACTGATGAACGAACTCGACAGCGGTGATCGGC 1136

QY      61  AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaIleGlnSerTyr 80
DB      1137  AACACGAAAGCCCTGCTGCTATGCTGGAATGCTGCTGGCGCAGGCATGCGCAAGCTAT 1196

QY      81  SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrp 100
DB      1197  TCCCAACGCACTCATCTCTTAAAGATATCGATATCGCGCGCGGAAACCAACCCGATTTG 1256

QY      101  IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
DB      1257  ATTTTGTCCACAAACAGTCGCGGCGGTGCTAGCTGATTTGTTGCCGCGGATTTG 1316

QY      121  LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrIleuHisLeu 140
DB      1317  AAGGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTTGACTTATCTGCACCTG 1376

```

```

QY      141  MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
DB      1377  ATGCGCTGTTTAATGCTGAGAGCGCAAGACGCGGCTATCGGTGACGAGCTAC 1436

QY      161  ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
DB      1437  CGCATGTCTCAATCGGCACACTGGGCACAAATAGCGCACTTGGCGGAGTCAATGCTGCGCTG 1496

QY      181  HisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSerAsnGluHis 200
DB      1497  CACGAGCCGCGCATTTCCGCGCTGCTGATTTTATCTTCAACCACTCCCAACGACAC 1556

QY      201  GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
DB      1557  GAATGGGCGCAACGCTGCGCGCGCGGACCGCTTTTTCGACAATTTCTACTATATTTTC 1616

QY      221  ProAspArgArgMetProAsnGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
DB      1617  CCGACCGCGGATGCGGACCAATAGCACCCCTTCCGCGGAAATCTTCCCGGACCG 1676

QY      241  HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsnSer 260
DB      1677  CACCGGCGGCTTCTCGCAACTGGAAGACGAGCGTGGGTGTGGACGACCTTCAATATCC 1736

QY      261  PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280
DB      1737  TTCCAAATGGGACTTGAATTAACAGCAACCCGCGGTATTTCCGCGCAATGCGCGGGAATG 1796

QY      281  LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300
DB      1797  CTGTTCTCTGCAACTTTGGCGCTTGACATCTCGGTATGATGCGTGGTTTATTTTGG 1856

QY      301  LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
DB      1857  AAACAAATGGGGACARGCTCGCAAAACCTCGCGCAGCGCGCACGCCCTCATCGCGCGGTC 1916

QY      321  AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
DB      1917  AATGCGGTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAATCCGAGGCAATGCTCCAC 1976

QY      341  ProAspGlnValValGlnTyrIleGlyGlnAspGlnCysGlnIleGlyTyrAsnProLeu 360
DB      1977  CCCGACCAAGTCGTCCAAATACATCGGCGCAGCAAGTCCCAATCGGTTTCAACCCCTG 2036

QY      361  GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380
DB      2037  CAAATGGCATTTGTGTGAACACCTTTGCCACGCGCGGAGTCAACCTGCTCCATCAGCGG 2096

QY      381  LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400
DB      2097  CTGACCTACGCGCAACACTGCGCGAGCATACCGCTGGGTCAACTACGTCGCGCAGCCAC 2156

QY      401  AspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420
DB      2157  GACGACATCGGCTGACGCTTTGCCGATGAAGACGCGGCATATCTGGGCATGAAGCGGCTAC 2216

QY      421  AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
DB      2217  GACCACCGCAATTCCTCAACCGCTTCTTCTGTCGTCGTCGCGCAGCTTCTGCTCGT 2276

QY      441  GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
DB      2277  GGGTACCGCTTCCAAATCAACCCCAAGCACAGCGGACGTCGCGTGTCACTGGTACGCGCGG 2336

QY      461  AlaLeuValGlyLeuAlaGlnAspProHisAlaValAspArgIleLysLeuLeuTyr 480
DB      2337  GCATTTGTCGGCTTGGCGCAAGACGATCCCAACCGCGCTTGACCGCATCAAACTCTTGTAC 2396

QY      481  SerIleAlaLeuSerThrGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
DB      2397  AGCATTCCTTTGATACCGCGGCTCTCGCGCTGATTTTACCTAGCGCAGGAATGGGTACG 2456

```

QY 501 LeuAsnAspAspTpsSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis 520
 Db 2457 CTCATATGAGGAGAGCTGTCGCAAGACCAATATAGAGGACAGACGCGTTGGCGGCAC 2516
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
 Db 2517 COTCGCGCTACACGAGCCCTGTACGCGCAACGCAACATCCGTCGACGCGCGG 2576
 QY 541 GlnIleFyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 Db 2577 CAAATCTATCAGGCTTGGCCCATATGATGTCGCGCCGCAAGCAATCCGCGCTTCGAC 2636
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580
 Db 2637 GCGCGCAGGCTGGTTACATTTCAACCAACCAACAGCACATCATCGCTACATCCGCAAC 2696
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
 Db 2697 AATGCGCTTTTGGCAATTCGGTAATTCACGCAATATCCGCAACCGTTACCGCGCATACC 2756
 QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
 Db 2757 CTGCAAGCCATGCCCTTCAGGCGGACGACCTCATCGTGGCAAACTGTGAGCTCGAT 2816
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
 Db 2817 CAGGATTTGACCTTCAGCCCTATCAGGTCTATCGTGGCTCGAAATCGCC 2864

RESULT 2

US-09-815-242-7764
 ; Sequence 7764, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, E. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7764
 ; LENGTH: 3303
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(3303)
 US-09-815-242-7764

Alignment Scores:

Pred. No.: 7,86e-56 Length: 3303

Score: 541.50 Matches: 165
 Percent Similarity: 41.60% Conservative: 90
 Best Local Similarity: 26.92% Mismatches: 223
 Query Match: 15.89% Indels: 135
 DB: 9 Gaps: 25
 US-09-843-007A-2 (1-636) x US-09-815-242-7764 (1-3303)
 QY 96 AsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyValCysTyr----- 112
 Db 31 AACACCCGCAATGGTACAAGGAC-----GGGTGATCTACAGGTCCAC 75
 QY 113 ValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLysAsp 125
 Db 76 GTGAAATCCCTTCTACGAGCGCCACCAACGATGGCATCGGACCTTCGCGGCTGATCGAG 135
 QY 126 LysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLys 145
 Db 136 AAGCTGACTACATCGCGGACCTCGCGGTGAACACTCTCTGGCTGTCTCCCTTTCTAC--- 192
 QY 146 CysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnPro 165
 Db 193 ---CCGTGCGCACCGCCGCGACGCGCTACGACATCGCCCGTACCGTGGCTGGCTGCACG 249
 QY 166 AlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIle 185
 Db 250 GACTACGCGACCTCGCGGACGCGCGGCTTTCATCGCGGAGGCCATCGACGCGGTCTG 309
 QY 186 SerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp----- 202
 Db 310 CGGTGATTACCGAGCTGGTGATCAACATACCTTCGACGACGATCCCTGGTTCATCGT 369
 QY 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222
 Db 370 GCCCGCACGCAAGAGGATCGCGCGCCGCGAC-----TACTACGCTCTGGTCGGAC 423
 QY 223 ArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisPro 242
 Db 424 -----AGCGACGAGAGAAATACCGAGGTACG---CGGATCATCTTCATCGAC--- 465
 QY 243 GlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThr----- 256
 Db 466 -----ACCGAGCAGTCCCACTGGACCTGGGACCCGCTAGCCCAACAGTAC 510
 QY 257 -----ThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPhe 273
 Db 511 TACTGGCACCGCTTCTATTCACCGACGCGACTGAACCTTCGACCAACCCGCGAGTCCGT 570
 QY 274 ArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMet 293
 Db 571 CGCGAGGTGCTCGGGGTGATCGCTACTGCTGACATGGGCGCTCGACGCGCTTGGGCTG 630
 QY 294 AspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGlnAsnLeuProGlnAla 313
 Db 631 GACCGGATTCCTTACCTGATCGAAGCGCGACCGCACCGACGAGAGAACTTCGCGGAGACC 690
 QY 314 HisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePhe 333
 Db 691 CACAGGTGCTCAAGCGCATCGCGCGGAGTGGACGCGGCACTATCCCGACCGCATGCTG 750
 QY 334 LysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly----- 349
 Db 751 CTGGCGAGGCCAATCAGTGGCGGGAAGACACCCGCGCTACTTCGCGCGGAGGATGGC 810
 QY 350 -----GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsn 367
 Db 811 GCGAGGGGCGAGCAATGCGCATGGCTTCCACTTCGCGTGTATGGCGCGCATGTACATG 870
 QY 368 ThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeu 387
 Db 871 GCCATCGCCCGAGGAGTGGCTATCCGATCACCGACATCTCTGGCCGACAGCCCGGACATC 930
 QY 388 ProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPhe 407


```
Db 931 CCGGCAATGGCAATGGCGATCTTCTGCGCAACACGAGAGCTCCCTCGAGATG 990
Qy 408 AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn 427
Db 991 GTACCGAC
Qy 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
Db 1018 -----TGGAAC 1023
Qy 448 ProSerThrGlyAspCysArgValSer-----GlyThrAlaAlaAlaValGly 464
Db 1024 CACTATGCCCGCGACCGCGCGCGCCCTCAACCTGGGGATCCCGCGCGCTCGCGCG 1083
Qy 465 LeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeu 484
Db 1084 CTGGTGGAGCGTGCACCGCGG-----CGCATCGAGCTGCTGCACAGCCTGTGTG 1134
Qy 485 SerThrGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr----- 500
Db 1135 TCGATGCCCGGCGACCGCGGACCTGTACTACGGCGAGAGATCGGCATGCGGCGACACATC 1194
Qy 501 ---LeuAsnAspAsp-----TriSerGlnAspSerAsn--- 511
Db 1195 TACCTCGCGACCGCGAGCGGCGGCGGACCGCGATGCTGCTGCGACCGCAACGCG 1254
Qy 512 -----LysSerAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu 528
Db 1255 GCGTTCTCCCGCGCGACCGCGGAGAGAGCTGGTG---CTCGCCCGGATTCGACCGCGTG 1311
Qy 529 Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly 540
Db 1312 TACGGCTACAGAGATCAAGCTGAGGCGGCGGCGGCGGACCGCATTCG----- 1362
Qy 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
Db 1363 ---CTGCTCAACTGATGCGCGCTGCTGCGCGTGGCGAGCAGATCTCTGCGTGGCGGCGGCGG 1419
Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIle----- 578
Db 1420 CGCGGAGCGCTGAAGATGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1479
Qy 579 -----ArgAsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrPro 593
Db 1480 TACGCGGAGCGGAGCGGAGGAGACAGATCTCTGCGTGGCGGCGGCGGCGGCGGCGG 1539
Qy 594 GlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis----- 609
Db 1540 CAGCGCGTG-----GAACCTGACCTCGCGGCGGCGGCGGCGGCGGCGGCGG 1581
Qy 610 -----AspLeuIleGlyLysThrVal-----SerLeuAsnGlnAspLeu 623
Db 1582 CGGCTGGAGATGATCGCGGCGATGCTGCTCCCGCGGATCGGCGGAGCTGACCTGCTG 1641
Qy 624 ThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
Db 1642 ACCCTGCGGCGCTACGCGCTTCTACTGCTTCTACCTGCGC 1680
```

RESULT 3

```
US-10-156-761-2790
; Sequence 2790, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIOKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2790
; LENGTH: 1716
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1716)
US-10-156-761-2790
```

```
Alignment Scores:
Pred. No.: 8,49e-55 Length: 1716
Score: 529.50 Matches: 166
Percent Similarity: 41.31% Conservative: 86
Best Local Similarity: 27.21% Mismatches: 233
Query Match: 15.54% Indels: 125
DB: 14 Gaps: 23
```

US-09-843-007A-2 (1-636) x US-10-156-761-2790 (1-1716)

```
Qy 93 AlaArgGluAsnAsnProAspTriPile----- 101
Db 46 GCCAAGGACCCGATCCGGAGTGTTCAAACGCCCGCTCTTCTACGAGGCTCTGTCGC 105
Qy 102 ---LeuSerAsnLysGlnValGlyIleValCysTyrValAspLeuPheAlaGlyAspLeu 120
Db 106 TCCTTCCAGGACAGCAACGCGGCGGCTGTC-----GGCGACCTG 144
Qy 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
Db 145 AAGGCTCTGACCGCCAGCTGACTATCTGCGTGGCTGGCGGCTGCTGCTGCTGCTG 204
Qy 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr 160
Db 205 CGCGCTTCTTCAAGTCCCGC-----CTGCGCGACGCGGCTACGAGCTCTCGGACTAC 258
Qy 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
Db 259 ACCGCGGTGTCGCCGAGTTCGTCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 318
Qy 181 HisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHis 200
Db 319 CACCAGCGCGCATCGCGGTGATCATCGACTTGTGTGTAACACACACACGAGCTGCAC 378
Qy 201 GluTrpAlaGln-----ArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyr 218
Db 379 CCGTGGTTCAGGAGTCCCGACGACCCCGCGGCTTACGCGAC-----TACTAC 432
Qy 219 IlePheProAspArgArgMetProAspGlnTyrAspThrLeuArgGluIlePhePro 238
Db 433 GTGTGGCGGAC-----GACGACAAGCAGTACACGAGGAGCGCGGCTATCTTCGTC 483
Qy 239 AspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpVal----- 254
Db 484 GAC-----ACCGAGGCTTCCAACTGACCTTCGACCCCGGTC 519
Qy 255 -----TriPThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269
Db 520 CGCAAGCAGTACTACTTGGCAGCGCTTCTTCTCCACGAGCGCGGCTCACTCACTACGAGAAC 579
Qy 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
Db 580 CGCGCGGTGCAGGAGAGATCGTCTCGCGCTTCTGGCTCGACTCGGCTCGGCTCGGAC 639
Qy 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309
Db 640 GCGTTCGCGCTCGAGCGGCTGCGGTACTGTGTACGAGGAGGAGGACCACTCGCGAGAAC 699
```

310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro 329
 700 CTGCGCGGACGACGAGTTCCTGAAGCGGGTGGCGAAGGAGATCGACGCACTATCCG 759
 330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349
 760 GACACGGTGTCTGCGCGAGGCGAACCAGTGGCGGAGGACGTGGTCTGACTACTTCGGC 819
 350 Gln-----AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeu 364
 820 GACTTCCTCCGCGGCGGCGACGAGTGCACATGGCGTTCCTATTTCCCGGTCATCCGCGG 879
 365 LeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArg 384
 880 ATCTTCATGGGGTGGCGGTGAGTCCGGTATCCGGTGTCCGAGATCTTGGCGAAGACG 939
 385 HisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHisAspAspIleGly 404
 940 CCGCGCATCCGTGAGTGCACGTGGGATCTCTCTGCGCAACCAACGACGAGCTGACC 999
 405 TrpThrPhe--AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArg 423
 1000 CTGGAGATGTGTCACGACGAGGAACGCGACTACATGTGG----- 1038
 424 GlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValPro 443
 1039 -----GCGAGTACGCGCAAGGATCCGCGG 1062
 444 PheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuVal 463
 1063 ATGCGGCGCCAAC-----ATCGCATCCCGCGCGCTCTGCGG 1098
 464 GlyLeuAlaGlnAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAla 483
 1099 CCGCTGCTGGACCAACGACGCG-----AACCAGATCGAGCTGTTCACCGCGCTGTG 1149
 484 LeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr----- 500
 1150 CTGTCGTGCCCGCTCCCGCATCTCTACTACGCGACGAGATGGCGATGGGGGACAAC 1209
 501 -----LeuAsnAspAspAsp-----TrpSerGlnAspSerAsn 511
 1210 ATCTGGCTCGTGACCGGACGCGGTGGCGACTCCGATCGAGTCGACGCGGACCGCAAC 1269
 512 -----LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu 528
 1270 GCGGGTTCTCGTCTCGACCCCGGGCGGTGTATTCGCCACGATCATGAGTCCCGGT 1329
 529 Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly 540
 1330 TACGGTACCAGTCCAGAACGTGGAGCGTCCGATGCTCGCGCTCTCTCG----- 1380
 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 1381 ---CTGCTGCATCGACCGCGCGATGATCGAGATCCGTAAAGCAAGAACCCCGCGTCTCGGC 1437
 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg--- 579
 1438 CTGCGTCTGTACACCGAACTCCAGTCTCTGNAACCGCGCGTCTCTCGGTTCCTCGGGAG 1497
 580 -----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu 591
 1498 GCCCCTCTGACCGGGGGGAACGGGGACGACCTGTGTGTGTGTCGTGCACACTCTCTCCGG 1557
 592 TyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis----- 609
 1558 TTCGCGAGGCC---ACGAGCTCGATCTCGGGCGGTTCACGCGCCGT---CATCCGTC 1611
 610 AspLeuIleGlyLysThrVal-----SerLeuAsnGlnAspLeuThrLeu 625
 1612 GAGCTGATCGCGGTGTCCGCTTCCCGGCCCATCGGGAACTCCCGTATCTGCTACACCTG 1671

```

Cy      626 GlnProTyGlnValMetTrpLeuGluIle 635
Db      1672 GCAGGCCACGGCTTCTACTGTTCCGGCTC 1701
      ::::
      |||
      ::::

RESULT 4
US-09-738-626-2527
; Sequence 2527, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCES: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2527
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2527

Alignment Scores:
Pred. No.: 9,11e-55 Length: 1794
Score: 529.50 Matches: 153
Percent Similarity: 46.32% Conservative: 105
Best Local Similarity: 27.47% Mismatches: 218
Query Match: 15.54% Indels: 81
DB: 10 Gaps: 22

US-09-843-007A-2 {1-636} X US-09-738-626-2527 {1-1794}

```

```

Alignment Scores:
Fred. No.:          9,11e-55           Length:         1794
Score:             529.50              Matches:        153
Percent Similarity: 46.32%            Conservative:   105
Best Local Similarity: 27.47%         Mismatches:    218
Query Match:       15.54%            Indels:        81
DB:                10               Gaps:         22

US-09-843-007A-2 (1-636) x US-09-738-626-2527 {1-1794}

Qy      118 GlyAspLeuIysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      223 GGATCGTTGAAGGCCCTGCACCGAAACTGGATTATCATCCAGTGCTCGCGTGATGCC 282

Qy      138 LeuHisLeuMetProLeuPheLysCysproGluGlyLysSerAspGlyGlyTyrAlaVal 157
      ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      283 ATTTCGGATCCCACCCTTTTANGATTCCCCA-----CTGCGCGACGGCGGTTCACGATATC 336

Qy      158 SerSerTyrArgaspValAsnProAlaLeuGlyThrlleGlyAspLeuArgGluValIle 177
      ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      337 CGCAACTTCCTGGTAATCCTCCCGAATTCGGCACCGTCGATGACTTCGTGGAACTCGTT 396

Qy      178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
      ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      397 GACCACGCCACCCCGCTGGCGCTCGGTGTATCCGACTTGGTCATGAATCACACCTCCC 456

Qy      198 AsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215
      ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      457 GACCAGCACGCATGTTCCAAGMA--TCCCGCGCGCACCCAAACCGGCCCTTACGGAGAT 513

Qy      216 PheTyrTyrllePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu 235
      ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      514 TTCATGTGTGAGCCGATGAT-----CCACCCCTGTATCAACGAAGACC---CGCATC 561

Qy      236 IlePheProaspGlnHisProGlyGlyPheSer---GlnLeuGluaspGlyArgTrpVal 254
      ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

```

QY      586  pheGlyAanPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro 605
Db      1606  GTCACAAATCATGAGCAAAATATCTCTCAGGCAATC-----TCGCTTGATTGGGTGAA 1656
QY      606  PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu 619
Db      1657  TTTCGACGACACACCCCTCGAGAGATGTCGGCGCGGACAGCTGTTCCCTTACCATTCGTGAA 1716
QY      620  AsnGln---AspLeuThrIleuGlnProTyrGlnValMetTrpLeuGluIle 635
Db      1717  CGGGAGTGGATGTCTCACTTTAGCCCTCACGGATTCCTTCGTGTGATCTC 1767

RESULT 5
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOL, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.: 1,5e-49 Length: 3309400
Score: 529.50 Matches: 153
Percent Similarity: 45.32% Conservative: 105
Best Local Similarity: 27.47% Mismatches: 218
Query Match: 15.54% Indels: 81
DB: 10 Gaps: 22

US-09-843-007A-2 (1-636) x US-09-738-626-1 (1-3309400)
QY      118  GiVAspLeuHyGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
Db      2438335  GGATCGTTGAAAGGGCTGACCGAAACATGGATTACCATCGCTCGCGTGGATTGC 2438394
QY      138  LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157
Db      2438395  AITTTGATCCACCGCTTTTATGATCCCA-----CTGCGGACGGCGTTACGATATC 2438448
QY      158  SerSerTyrArgAspValasnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db      2438449  CGCACTTCGGTGAATCTCGCCGAATTCGGCACCGTCGATGACTTCGTGGAACTCGTT 2438508
QY      178  AlaAlaLeuHisGluAlaGlyIleSerAlaValAlaPheAsnHisThrSer 197
Db      2438509  GACCACGCCACCGCGCTGGCTTATCACCAGACTTGGTCATGAATCACACTCC 2438568
QY      198  AsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215

```

```

Db 2438569 GACCAGCAGCATGGTTCACAGAA...TCCCGCGCGACCAACCGCGCCCTACGGAGAT 2438625
Qy 216 PheTyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu 235
Db 2438626 TTCATGTGTGAGCGAGATGAT...CCACCCCTGTACACAGAGCC...CGCATC 2438673
Qy 236 IlePheProAspGlnHisProGlyGlyPheSer...GlnLeuGluAspGlyArgTyrVal 254
Db 2438674 ATCTTGTAGATACAGAGATCCAGACTGACCTATGATCGGTGGTGGCCAGTACTTC 2438733
Qy 255 TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArg 274
Db 2438734 TGGCACCGCTTCTCTCCACCAACAGACAGCTCACTACGACAAACCCCGCAGCTCCAGAG 2438793
Qy 275 AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp 294
Db 2438794 GCCATGCTAGATGCTCTCGCTTCTGCTGACCTTGGACCTTGATGTTTCCGACTAGT 2438853
Qy 295 AlaValAlaPheIleTrpPheGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis 314
Db 2438854 GCCGTCTCTTATCTTTTGAACGCGAAGCCACCAACCGCGAAACCTCAAGAAACCCAC 2438913
Qy 315 AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys 334
Db 2438914 GATTTCTCAACTGTGCTCTGTCTGATGAGAGGAATACCCCGCGCGAATCTGCTGCTC 2438973
Qy 335 SerGluAlaIleValHisProAspGlnValValGlnTyrIleGlyGln...----- 350
Db 2438974 GCAGAGGCCAACCAATGCCCAAGATGTGTGCAATACTTTCGGTGAAAGACACAAAGCC 2439033
Qy 351 AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrPheAsnThrLeuAla 370
Db 2439034 GATGAATCCCATGCGGCTTCCACTTCCCTTTGATGCGCGCATCTTCATCGGAGTTCGC 2439093
Qy 371 ThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlnHis 390
Db 2439094 CAAGTTCCAGCCACCCCATGAGATCTGCGCCACACCCCGGAGATTCACAGACT 2439153
Qy 391 ThrAlaTrpValAsnTyrValArgSerHisAspIleGlyTyrThrPhe...-AlaAsp 409
Db 2439154 GCCCAATGGGTATTTTCTCGTGAATCATGATGATGATGATGATGATGATGATGATGAT 2439213
Qy 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn...- 427
Db 2439214 CAGGACCCGACTACATG...TACTCCCAATCGCTCCGACCACT 2439255
Qy 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
Db 2439256 CGCATCGCGCCAA...-----GTAGGATCCGCGCGCCCTTCCCACTCTTGA 2439306
Qy 468 AspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly 487
Db 2439307 GCGCAGCCG...-----AACGAGTGAACCTCTTCAAGGTTTCTGCTCTCTACCT 2439357
Qy 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr...-----LeuAsn 502
Db 2439358 GGCCTCACCGGTGTGTATTACGGTATGAAATGGCATGGCGCAATATCTGCTCCAC 2439417
Qy 503 AspAspAsp...-----TrpSerGlnAspSerAsn...----- 511
Db 2439418 GACCGCAGGAGTTCGCGACCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2439477
Qy 512 LysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeu...-TyrAla 530
Db 2439478 AAAGCTGATCTGTAACGCGCTGTACCTTCCAGCGATCCCAAAATGATCAATACGCTACGCC 2439537
Qy 531 GlnArgAsnAspProSer...-----ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis 548

```

RESULT 6

```

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187115)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

```

Alignment Scores:
Pred. No.: 7,44e-49 Length: 9025608
Score: 529,50 Matches: 166
Percent Similarity: 41,31% Conservative: 86
Best Local Similarity: 27,21% Mismatches: 233
Query Match: 15,54% Indels: 125
DB: 14 Gaps: 23
US-09-843-007A-2 (1-636) x US-10-156-761-1 (1-9025608)

```

```

Qy 93 AlaArgGluAsnAsnProAspTrpIle...----- 101
Db 3441583 GCCAAGACCGCATCGGAGTGTTCAAACGGCGCGCTTCTTACGAGGTCTGTCGCCG 3441642
Qy 102 ---LeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
Db 3441643 TCCTTCAGGACACACGCGCGCGGTGTC...-----GGCGACCTG 3441681
Qy 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140

```

DB 3441682 AAGGGCTGACCGCAAGCTGGACTATCTGAGTGGGTGGCGGTGAGTGGCTGTGGCTG 3441741
QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160
DB 3441742 CCGCGTCTTCAAGTCCCC-----CTGCGGACGCGGCTAGACGCTTCCGACTAC 3441795
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
DB 3441796 ACCGGGTGCTGCCCGAGTTCGGTGACTCCGCGCACTTCGTCGAGTTCGTGGACGCGGCC 3441855
QY 181 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200
DB 3441856 CACACGCGCGGTGGCTGATCATCGATCTTGGTGAACCAACACCGGACGACTCGAC 3441915
QY 201 GluTrpAlaGln-----ArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyr 218
DB 3441916 CCGTGTCTCCAGGAGTCCCGAGCAACCCGACGCGCCCTACGCGGAC-----TACTAC 3441969
QY 219 IlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhePro 238
DB 3441970 GTTGGGCGGAC-----GACGACAGCAGTACACGAGGAGCGCGGATCATCTTGTCT 3442020
QY 239 AspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrVal----- 254
DB 3442021 GAC-----ACCGAGGCTCCAACTGGACCTTCGACCGCGTC 3442056
QY 255 -----TrpThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsn 269
DB 3442057 CGCAAGCAGTACTACTGCGACCGCTTCTTCCACGCGCGGACCTCACTAGAGAAC 3442116
QY 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
DB 3442117 CCGCGGTGCGAGGAGAGTCTCTCGCGCTCGCGTCTGCTGCGACCTCGCATCGAC 3442176
QY 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309
DB 3442177 GGTTCGCGCTCGACGCGGTGCGTACCTGTACACGAGGAGGACCACTGCGAGAAC 3442236
QY 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro 329
DB 3442237 CTGCGCGGACGACAGTCTCTGAAGCGGTGCGCAGGAGATCGACACGACTATCCG 3442296
QY 330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349
DB 3442297 GACACGGTGTCTCGCGAGGCGGACCAAGTGGCGGAGGAGTGTGCTGACTACTTCGCG 3442356
QY 350 Gln-----AspGluCysGlnIleGlyTyrAsnProLeuGluMetAlaLeu 364
DB 3442357 GACTTCCCTCGCGCGGCGGACGAGTCCACATCGGTTCCTATTTCCCGGTATCCCGCGG 3442416
QY 365 LeuTrpAsnThrLeuAlaThrArgGluValAlaAsnLeuHisGlnAlaLeuThrTyrArg 384
DB 3442417 ATCTTCATGGCGGTGCGGCGTGGTATCGGTATCGGTGCGGATCTCGCGAAGACG 3442476
QY 385 HisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspIleGly 404
DB 3442477 CCGCGGATCCCGTCCGAGTGGCGGATCTTCTTGGGCAACCAACGACGAGTGCACC 3442536
QY 405 TrpThrPhe-----AlaAspGluAspAlaValTyrLeuGlyIleSerGlyTyrAspHisArg 423
DB 3442537 CTGGAGATGGTCAACGACGAGGACGCGACTACATGTG----- 3442575
QY 424 GlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValPro 443
DB 3442576 -----GCGAGTACGCGCAAGATCCGCGG 3442599
QY 444 PheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuVal 463
DB 3442600 ATGCGGGCAAC-----ATCGGATCCGCGCGGTCTGCG 3442635
QY 464 GlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAla 483
DB 3442636 CGCTGCTGGACACGACGCG-----AACGATGAGCTGTTCACGCGGTCTGTG 3442686

QY 484 IeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspIleValGlyThr----- 500
DB 3442687 CTGTGCTGCCCGCTCGCGGATCTCTACTACGCGGACGAGATCGCATGGGCGGCAAC 3442746
QY 501 -----LeuAsnAspAsp-----TrpSerGlnAspSerAsn 511
DB 3442747 ATCTGCTCGGTACCGGACGCGGTGCGGACTCCCATCATGATGAGTCCGCGGAC 3442806
QY 512 -----LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu 528
DB 3442807 GCGGGTTCCTGCTCGGACCGCGCGGTCTGTATCTGCCCACGATCATGATCGGTGTC 3442866
QY 529 Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly 540
DB 3442867 TAGCGTACAGTCCAGAACGCGGAGGCTCGATGTCGCGCGTCTCG----- 3442917
QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
DB 3442918 ---CTGCTGCACTGCGCGCGGATGATCGATCGTAAAGCAGAACCGCGGTTCGCG 3442974
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg--- 579
DB 3442975 CTGCGCTGTACACCGAACTCCAGTCTCGAACCCCGCGCTCTCGGTTCCTCGCGGAG 3443034
QY 580 -----AsnAsnAlaLeuAlaPheGlyAsnPheSerGlu 591
DB 3443035 GCCCTCGACCGGCGGAGACGCGGAGCTGTGCTGCGTGCACACTTCTCCCGG 3443094
QY 592 TyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis----- 609
DB 3443095 TTGCGGACGCGC---ACGAGCTGGATCTCGCGGCTTCAGCGCGCT---CATCGGTC 3443148
QY 610 AspLeuIleGlyGlyLysThrVal-----SerLeuAsnGlnAspLeuThrLeu 625
DB 3443149 GAGCTGATCGCGGTGTCCTCCCGGCACTCCCGTATCTGCTGACCTG 3443208
QY 626 GlnProTyrGlnValMetTrpLeuGluIle 635
DB 3443209 GCAGGCCACGCTTCTACTGTTCCGCGCTC 3443238

RESULT 7
US-10-156-761-7374
; Sequence 7374, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7374
; LENGTH: 1749
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1749)
US-10-156-761-7374
Alignment Scores: 1.49e-52 Length: 1749
Pred. No.:

Score: 511.50 Matches: 163
 Percent Similarity: 40.98% Conservative: 87
 Best Local Similarity: 26.72% Mismatches: 235
 Query Match: 15.01% Indels: 125
 DB: 14 Gaps: 23

US-09-843-007a-2 (1-636) x US-10-156-761-7374 (1-1749)

Qy 93 AlaArgGluAsnProAspTrrile----- 101
 Db 46 GCCAAGACCGCATCGGAGTGTTCAAAGCGCCGCTTCTTACGAGGTCTCGTCCGC 105
 Qy 102 --LeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
 Db 106 TCCTTCAGGACACGACCGCGCGTCTC-----GGGACCTG 144
 Qy 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
 Db 145 AAGGCGCTGACCGCAAGCTGAGCTATCTGAGTGGCTGGCGTGGAGTCTGCTGCTG 204
 Qy 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr 160
 Db 205 CCGCGCTTCTCAAGTCCCTC-----CTGCGGACGCGGCTACGAGCTCTCGACTAC 258
 Qy 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
 Db 259 ACCGGGTGCTCCCGAGTCCGTCACCTGCGCGACITGTCGAGTTCGTGCAACGCGCC 318
 Qy 181 HisGluAlaGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSerAsnGluHis 200
 Db 319 CACCAGCGCGCATGCGGTGATCATGACTTCGTGATGAAACACACACGAGTGCCTGAT 378
 Qy 201 GluTrpAlaGln-----ArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyr 218
 Db 379 CCGTGTTCAGGAGTCCCGCAGCAACCCCGAGCGCCCTACGCGAC-----TACTAC 432
 Qy 219 IlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhePro 238
 Db 433 GTGTGGGCGGAC-----GAGCACAAGCAGTACCAGGACGCGCGCATCATCTGCTC 483
 Qy 239 AspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrVal----- 254
 Db 484 GAC-----ACCGAGCGCTCCAACTGAGCTAGCAGCCCGCTC 519
 Qy 255 -----TrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269
 Db 520 CGCAAGCAGTACTACTGCGACCGCTTCTTCTCCACAGCCGACCTCAACTACGAGAGT 579
 Qy 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
 Db 580 GCCCGCTCCAGGAGGAGATCTGCGCGCGCTGCGGTTCGTGCTCGATCTGGGCATCGAC 639
 Qy 290 IleLeuArgMetAspAlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsn 309
 Db 640 GGCTTCAGCTGGACCGCTCCCTACTGTACACGAGAGGAGGAGGAGCACTGGGAGAC 699
 Qy 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAlaValMetArgIleAlaAlaPro 329
 Db 700 CTGCGCGGACGACGAGTCTCTCAAGCGGTGGCGCAGGAGATCGACACGACTATCCG 759
 Qy 330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGly 349
 Db 760 GACAGGTGCTGCTCGCGAGGCGAACCAGTGGCGGAGGAGCTGTGCTGACTTTCGCG 819
 Qy 350 Gln-----AspGlyCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeu 364
 Db 820 GACTTCCCTCCGCGCGGCGACGAGTGCACATGCGGTTCATTTCCCGGTCCATCGCGG 879
 Qy 365 LeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArg 384
 Db 880 ATCTTCATGGCGGCGGCGGTGAGTCCGCGTATCCGCTGCGGAGATCTTCGCGAAGCG 939
 Qy 385 HisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspIleGly 404

RESULT 8

US-08-781-986A-53/c
 ; Sequence 53, Application US/08781986A
 ; Publication No US2003005436A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Charles Kunsch
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 5255
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA

Db 940 CCGCGCATCCGTCGAGCTGCCAGTGGGCGATCTTCTGCGCAACACGACGAGCTGACC 999
 Qy 405 TrpThrPhe-----AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArg 423
 Db 1000 CTGAGATGCTCACCGACGAGGAACGCGACTACATGTGG----- 1038
 Qy 424 GlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValPro 443
 Db 1039 -----GCGAGTAGTACGCCAAGGATCCCGCG 1062
 Qy 444 PheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuVal 463
 Db 1063 ATCGGGGCCAAC-----ATCGGCATCCCGCGCTCTGCG 1098
 Qy 464 GlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAla 483
 Db 1099 CCGCTGTGTGACAAACGACCGC-----AACAGATCGAGCTGTTCACCGCGCTGCTG 1149
 Qy 484 LeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr----- 500
 Db 1150 CTGTGCTGCCGCGCTCGCGATCTCTACTACGCGACGAGATCGGATGGGGGACAAC 1209
 Qy 501 -----LeuAsnAspAspAsp-----TrpSerGlnAspSerAsn 511
 Db 1210 ATCTGGCTCGGTACCGGACGCGGTGCGCACCGCATGCGAGTGGACGCGGACCGCAAC 1269
 Qy 512 -----LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu 528
 Db 1270 GCGGTTTCTGCTCTGCGACCGCGGCGTCTGTATCTGCCACGATCATGATTCGGTC 1329
 Qy 529 Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly 540
 Db 1330 TAGCGGTACCGACTCAGAACGCTGGAGCGCTCGATGCTGCTGCGCTCTCTCG----- 1380
 Qy 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
 Db 1381 ----CTGCTGCACCTGACCGCGCGAGTATCGAGATCGTAAGCAGAACCGCGCTGCGC 1437
 Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg--- 579
 Db 1438 CTGCGCTCGTACACCGAACTCCAGTCTCGCAACCCCGCGCTCTCTGCTGCTGCGGAG 1497
 Qy 580 -----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu 591
 Db 1498 GCCCCTTCGACCGGGGGAACCGGGGACGACCTGCTGCTGTGCTGCGACAACTTCTCCCG 1557
 Qy 592 TyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis----- 609
 Db 1558 TTCCGCGAGCCC---ACGAGCTGAGTCTGCGGCGCTTACGCGCGCGT---CATCCGCTC 1611
 Qy 610 AspLeuIleGlyLysThrVal-----SerLeuAsnGlnAspLeuThrLeu 625
 Db 1612 GAGTGTATCGCGGTGTCCGCTTCCCGGCGCATCGGGGAATCCCGTATCTGCTGACCCCTG 1671
 Qy 626 GlnProTyrGlnValMetTrpLeuGluIle 635
 Db 1672 GCAGGCCACCGGCTTCTACTGTTCCGCGCTC 1701

ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,986A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Benson, Bob
 REGISTRATION NUMBER: 30,446
 REFERENCE/DOCKET NUMBER: PB248PP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 53:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 16592 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-781-986A-53

Alignment Scores:
 Pred. No.: 1,736-29 Length: 16592
 Score: 338.00 Matches: 158
 Percent Similarity: 36.36% Conservative: 106
 Best Local Similarity: 21.76% Mismatches: 230
 Query Match: 9.92% Indels: 232
 DB: 8 Gaps: 34

US-09-843-007a-2 (1-636) x US-08-781-986A-53 (1-16592)

QY 6 GlnGlnValGly---LeuIleLeuGlnTyrLeuLysThrArg-----IleLeuAsp 21
 DB 4830 CAACAGCGTGGTGCAGCAATTTAAATATATTGAAACAAACAGCAGGATTTATTGAA 4771
 QY 22 IleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTyrArgGlnPheSer 41
 DB 4770 CTCGTCATTATAGATACAGATTAATAAATAAGAAATCAACACAGCGATGACCAAGGAGT 4711
 QY 42 ArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGlyAsn 61
 DB 4710 TTTAAATGAATACCAATGGTGGAA-----4684
 QY 62 AsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTyrGlnSerTyrSer 81
 DB 4683 ---GAAGCAGTA-----GCATATCAAGTATATCCA 4657
 QY 82 GlnArgAsnSerSerLeuLysAspIleAlaArgGluAsnProtsPyrIle 101
 DB 4656 AGA-----AGTTTAAATGAT-----AGTAATCACAGTGGTATT 4624
 QY 102 LeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeuLys 121
 DB 4623 -----GGGGAATTACCT 4612
 QY 122 GlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMet 141
 DB 4611 GGAATGATTGATTAATGAGCTACTTAAAGATTTCGGTATCGATGCTATTGGCTCAGT 4552
 QY 142 ProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyrArg 161
 DB 4551 CCAATGTTTAAATCACCT-----AATGATGACAAATGGTTATGATATTAGTACCTACCAA 4498
 QY 162 AspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeuHis 181
 DB 4497 GAGATTATCGATGAATTGGACGATGGAAGAGCTTTGATCGTTTATTAAAGGTGTTTAT 4438

QY 182 GluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGlu 201
 DB 4437 GATAGAGCGCATGAAGCTTTATTAGATTGTTAAATCATACATCTGATGACATCT 4378
 QY 202 Trp-----AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrIle 219
 DB 4377 TGGTTTATAGATCCAAATCTAGTAAAGACATATCCCAACAGTGAT-----TGGTACATT 4324
 QY 220 PheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAsp 239
 DB 4323 TGGCAAGATCCAAAG---CCAGAT-----GGC 4300
 QY 240 GlnHisProGlyGlyPheSerGlnLeuGluAspGly-----251
 DB 4299 TCTGAACCTTAACAACTGGGAAAGTATATTATGATCTACATGCGAATATGCTAAT 4240
 QY 252 -----ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269
 DB 4239 ACTGACCAATATTATTTCATTTTTCAGTAAACAAACCTGATTGAAATGGGTTAAT 4180
 QY 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaLeuGlyValAsp 289
 DB 4179 CCGAAGTTAGACATGCTGTATTGAAATGATGACTGGTGGTTGATAAAGCATGAT 4120
 QY 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309
 DB 4119 CGATTAGATGATGCAATTACGCATATTAAAGAGACCTTTGAAGCGGTGACTTACCT 4060
 QY 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArg-----325
 DB 4059 GTACCTGAGGTAAACATATGCCCGACGATTGATGATGATGATGATGATGATGATGAT 4000
 QY 326 -----IleAlaAlaProAlaValPhePheLys 334
 DB 3999 CMAACTGGTTACAGAGATGAAAGATCGCTCATTAAGTAAGTATGATGATGATGATGAT 3940
 QY 335 SerGluAlaIle---ValHisProAspGlnValValGlnTyrIleGlyGlnAspGluCys 353
 DB 3939 GGTGACGGAATGGTGTAAAGCCCTGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 3880
 QY 354 GlnIleGlyTyrAsnProLeu-----GlnMetAlaLeuLeuTrpAsnThrLeuAla 370
 DB 3879 AAA-----TTTAAATGATATTCCCAATTGCAATTTGGGACTGTGGAATAGTGGTAT 3826
 QY 371 ThrArg---GluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389
 DB 3825 TCTCAGCTTGTATGATTAATTCGTATAATCTGTATTAAATAGATGGCAAAACCACTTGA 3766
 QY 390 HisThrAlaTrp---ValAsnTyrValArgSerHisAspAspIleGly-----TrpThr 406
 DB 3765 AATAAGGTTGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3706
 QY 407 PheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeu 426
 DB 3705 TGGGGTGACGATGACAAAGTATTGG-----TATGAA-----3676
 QY 427 AsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyr 446
 DB 3676 -----3676
 QY 447 AsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAla 466
 DB 3675 -----TCAGCAACAAAGTCATGACGCTGTTATTCTTCTG 3643
 QY 467 GlnAspAspProHisAlaValAspArgIleLysLeuLysSerIleAlaLeuSerThr 486
 DB 3642 CAACAA-----3637
 QY 487 GlyGlyLeuProIleTyrLeuGlyAspGluValGly-----499
 DB 3636 ---GGTACGCCATTCAATTATCAAGGTCAAGAAATGGTATGACCAATATCCATTTCAA 3580

QY 500 -----ThrLeuAsnAsp-----AspAspTTPSer----- 507
 Db 3579 AGTATTGAAACGTTTAAACGATGCTGCTTTAAATGACTATCAATATGTAAGTCAAAAGCTCAA 3520
 QY 508 -----GlnAspSerAsnLysSerAspAspSer 516
 Db 3519 GGTGGAGATGTAGACGCTTTACTTTCGAAATATAAAGATGAGAACCGA---GATAATCT 3463
 QY 517 Arg-----TtAlaHis----- 520
 Db 3462 CGCACACCAATGCAATGGATGATACGTTAAATGGAGGATTTACAAATGGTGAACCGTGG 3403
 QY 521 -----ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThr 537
 Db 3402 TTCACAGGAATCCGAATTAATAAACTATCAATGTGTCACAACTATTAGAAGATGAGCAT 3343
 QY 538 AlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnPro 557
 Db 3342 TCAGTATTCAATTTTAAAGATTTA-----ATTCAATTAAAGAAAGTCTAATCAT 3292
 QY 558 ArgPheAspGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyr 577
 Db 3291 GTATACGTATATGTCATTTGATTTAGTACATGCTGAAATTCACAAATTTTTCGCTAC 3232
 QY 578 IleArg-----AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGln 594
 Db 3231 ACAGAAACATTAATGAAGCAAGTCTTATTGTAGGTAATCTTACTAACCACGAGCA 3172
 QY 595 ThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGly 614
 Db 3171 GAATTAAT-----GTACCAAT-----GATTAAGCAATGA 3139
 QY 615 LysThrValSerLeuAsnGlnAspLeu-----ThrLeuGlnProTyrGln 629
 Db 3138 GAACTAAGCTATTATTAATGATGCAAGTAAATTAACAGTTACGTCCTATATGAA 3079
 QY 630 ValMetTTPLeuGluIle 635
 Db 3078 GCATGTGTTATCGAACTA 3061

RESULT 9

US-10-081-872-153
 ; Sequence 153, Application US/10081872
 ; Publication No. US20030125534N1
 ; GENERAL INFORMATION:
 ; APPLICANT: Callen, Walter
 ; APPLICANT: Richardson, Toby
 ; APPLICANT: Frey, Gerhard
 ; APPLICANT: Short, Jay M.
 ; APPLICANT: Mathur, Eric J.
 ; APPLICANT: Gray, Kevin A.
 ; APPLICANT: Kerovuo, Janne S.
 ; APPLICANT: Slupski, Malgorzata
 ; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
 ; TITLE OF INVENTION: AND METHODS OF USE THEREOF
 ; FILE REFERENCE: 09010-108001
 ; CURRENT APPLICATION NUMBER: US/10/081,872
 ; PRIOR FILING DATE: 2002-02-21
 ; PRIOR APPLICATION NUMBER: US 60/270,495
 ; PRIOR FILING DATE: 2001-02-21
 ; PRIOR APPLICATION NUMBER: US 60/270,496
 ; PRIOR FILING DATE: 2001-02-21
 ; PRIOR APPLICATION NUMBER: US 60/291,122
 ; PRIOR FILING DATE: 2001-05-14
 ; NUMBER OF SEQ ID NOS: 321
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 153
 ; LENGTH: 1620
 ; TYPE: DNA
 ; ORGANISM: Environmental
 US-10-081-872-153

Alignment Scores:

Pred. No.: 3,11e-30 Length: 1620
 Score: 331.00 Matches: 134
 Percent Similarity: 36.63% Conservative: 77
 Best Local Similarity: 23.26% Mismatches: 237
 Query Match: 9.71% Indels: 129
 DB: 14 Gaps: 20
 US-09-843-007A-2 (1-636) x US-10-081-872-153 (1-1620)
 QY 78 GlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAlaArgGluAsnAsn 97
 Db 110 CAATATACAGCAGAGAACATTAATTAATAGGTCATAAAG-GCGCAACAGGAAT 168
 QY 98 ProAspTTPLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPhe--- 116
 Db 169 AAAGACCGAATTTTATGAA-----CTGTATGTTAATCTTTTAT 210
 QY 117 -----AlaGlyAspLeuLysGlyLeuLysAspIleProTyrPhe 130
 Db 211 GATACTGATAGCAATGACATGCTGATTTAAAGGGGTCAAAAGAACTTGAATTTA 270
 QY 131 -----GlnGluLeuGlyLeuThrTyrLeuHisLeuMetPro 142
 Db 271 AATGATGGAAATCCAGAACAAATATATGATCTCAATTAACCGTATCTCGATGATCC 330
 QY 143 LeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAsp 162
 Db 331 ATTAACACACCTCTCT-----AGTTATCAAAATATGATGATTAACAGATTACTATA 381
 QY 163 ValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeuHisGlu 182
 Db 382 ATCGATCCTCAGTATGGAAGTTTACAGATTTCCGTGAACCTAACACAGAACGCGATA 441
 QY 183 AlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGluTyr 202
 Db 442 CGCAACGTAAGGTAGTAGTATAGATCTTGTTATTAATCATACAGCAGTACCTCTGG 501
 QY 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222
 Db 502 TTTGTGATGCAATTAATAAATAAAGACAGTAAGTATCGAGATTACTATATTTGGCTG 561
 QY 223 ArgArgMetProAspGlnTyrAspAlgThrLeuArgGluIlePheProAspGlnHisPro 242
 Db 562 AAAATACAGAC-----TTAATGAAAGGCGCCA 591
 QY 243 GlyGly-----PheSerGlnLeuGluAspGlyArgTyrValThrThrPheAsn 259
 Db 592 TGGGTCAACACAGTATGCGACAAAGCGTCGACGAGAGTATTTCTACGCAACGTTCTGG 651
 QY 260 SerPheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGlu 279
 Db 652 GAAGGATGCGGACTTAACCTATGACACCCCTAAAGTAGAGAGAAATGATTAATATC 711
 QY 280 MetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIle 299
 Db 712 GGGAAATTTGGCTCAACAGGAGCTGATGGCTTTGCTAGATGCGCATGCACATC 771
 QY 300 TrpLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAla 319
 Db 772 TTTAAA-----GGCAAAACCTCTGAAGGAGCAAGAAATAATTGAATGTCGAATGAA 825
 QY 320 PheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleVal 339
 Db 826 TTCGCGACGCGATGAGAGAAACGAATCCAAATACGTATCTAGTGTGTAATATGGAT 885
 QY 340 HisProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnPro 359
 Db 886 CAACCAAGATGATGCTCCGTAT----- 909
 QY 360 LeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGln 379
 Db 910 -----TATCAA 915

Qy 380 AlaLeuThrTyArgHisAsnLeuProGluHisThrAlaTrpValAsnTy-ValArgSer 399
Db 916 TCGTTAGATTCTACATTAACTTCGACTAGCATATAAAATCGTTAAATCGTTAAATAAT 975
Qy 400 HisAspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyLeuGlyIleSerGly 419
Db 976 GGTACTGATCAAGG-----GTAGCCCGCGCAGCT 1005
Qy 420 TyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAla 439
Db 1006 GTTGCAACGGATGATTATATAAACATATAATCAATATAATGATGGAACGTTTATA 1065
Qy 440 ArgGly-----ValProPheGlnTyAsnProSerThrGlyAspCysArg 454
Db 1066 ACGAATCATGACCAAAATCGTGTAAATGAGTGAATTAAT-----GGTGAT----- 1110
Qy 455 ValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAsp 474
Db 1111 -----GTAAAC 1116
Qy 475 ArgIleLeuLeuLeuTySerIleAlaLeuSerThrGlyLeuProLeuIleTyLeu 494
Db 1117 AAAGCAAAATCAGCAGCCTTATTCTGTGACACTCCCTCGTGAATCCGTTCAATTATAT 1176
Qy 495 GlyAspGluValGlyThrLeuAsnAsp-----AspAsp----- 505
Db 1177 GCGAGAAATCGGCATGACAGCCCAAAACAGATGAGTGAATCGTGAGCCCTTCGCT 1236
Qy 506 TrpSerGlnAspSerAsnLysSerAspSerArgTrpAlaHisArgProArgTyTrpAsn 525
Db 1237 TGGTATGAAGAT---GATAAAGAGGTCAAACGAGCTGG---GAGACTCCAGTATATAAC 1290
Qy 526 -----GluAlaLeuTyAlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIle 542
Db 1291 ATTGATCATATAGTGTTCAGTTGAGCACAAGATATAACAAAGCTTCT-----CTT 1344
Qy 543 TyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGly 562
Db 1345 CTAAGCCATTATCGTAAATGATTCGTGTCTGACACACGATGAACCTTGTCAAAGST 1404
Qy 563 ArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyIleArg----- 579
Db 1405 AATTTAGAACCTTATTCTGTCAATTAATTCACAGGTGTGTGCTATTAATCGTACGATATAA 1464
Qy 580 AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyTrpGlnThrValThrAlaHis 599
Db 1465 AATAATCAATTCAGTGTACCATATATTTTCAGACAAAGCCGTTACATTAACCTTTCA 1524
Qy 600 ThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLys 615
Db 1525 AACAAAGGAAACTGATTTTCTAGTGA-----TTAGGAGCAAAA 1566

RESULT 10

US-09-815-242-8494
; Sequence 8494, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8494
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1650)
US-09-815-242-8494
Alignment Scores:
Pred. No.: 6.54e-30 Length: 1650
Score: 328.50 Matches: 133
Percent Similarity: 37.38% Conservative: 95
Best Local Similarity: 21.80% Mismatches: 201
Query Match: 9.64% Indels: 181
DB: 9 Gaps: 27
US-09-843-007a-2 (1-636) x US-09-815-242-8494 (1-1650)
Qy 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyPheGlnGluLeuGlyLeuThrTy 137
Db 82 GGGGATTTACTGGAATGATTGATAAATCGACTACTTAAAGATTTAGGTATCGATGTC 141
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyAlaVal 157
Db 142 ATTTGGCTCAGTCCAAATGTTTAAATCACCT-----AATCATGATAATGTTATGATAT 195
Qy 158 SerSerTyArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 196 AGTGACTACCAAGAGATTATGGATGAATTTGGACGATGGAAGACTTTGATCGTTTATTA 255
Qy 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
Db 256 AAGGTCTTCATGATAGAGGTATGAAGCTTATTTAGATTTAGTTGTGAATCATACATCT 315
Qy 198 AsnGluHisGluTrp-----AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsn 215
Db 316 GATGAACATCTCTTGTTTATAGAAATCCAAATCTAGTAAAGACAAATCCCAACCGTAT--- 372
Qy 216 PheTyTrpIlePheProAspArgArgMetProAspGlnTyAspArgThrLeuArgGlu 235
Db 373 ---TGGTACATTTGGCAAGATCCAAAG---CCAGAT----- 402
Qy 236 IlePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGly----- 251
Db 403 -----GGCTCTGAACCTTAACACTGGGAAAGTATATTTAATGATCTACATCGGAA 453
Qy 252 -----ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeu 265
Db 454 TATGATCTTAATCTAGACATATATTTCCATTTATTCAGTAAACCAACACCTGATTG 513
Qy 266 AsnTySerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsn 285
Db 514 AATTGGGGTAAATCCGGAAGTTAGAGATGCTGTATTCCGAATGATGAATCGTGTGTTGAT 573
Qy 286 LeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThr 305
Db 574 AAAGGCATTGATGATTTAGATAGATGCAATTTACCATATTTAGAGAGACGTTTGAACG 633
Qy 306 SerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArg 325

```

Db      1522 AACCCAGACGAGCAATTAAC-----GTACCATT-----GAT 1555
Qy      611 LeuilecglylysthrValSerLeuAsnGlnAspLeu-----ThrLeu 625
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1555 GtaAGTCATGCGAGAGTGAAGCTATTTAATTATGATGCGCAAGCTTAATTTAAACACAGTTG 1614
Qy      626 GlnProTyrClnValMetTrpIleuGluIle 635
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      1615 CGTCCCATATGAAGCATGTGTTATCGAATTA 1644

RESULT 11
US-09-815-242-4455
; Sequence 4455, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCES: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4455
; LENGTH: 1638
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4455

Alignment Scores:
Pred. No.: 2,69e-29 Length: 1638
Score: 323.50 Matches: 132
Percent Similarity: 37.25% Conservative: 93
Best Local Similarity: 21.85% Mismatches: 198
Query Match: 9.49% Indels: 181
DB: 9 Gaps: 27

US-09-843-007A-2 (1-636) x US-09-815-242-4455 (1-1638)
Qy      118 GlyAspLeuLyScglyLeuLyAspLysIleProTyrPheIngluLeuGlyLeuThrTyr 137
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      82 GGGGATTACCTGGATGATGATTAATTGGACTACTTAAAGATTTAGGTATCGATGTC 141
Qy      138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      142 AATTGGCTCAGTCCCAATGTTTAATCACCT-----AATGATGATTAATGTTTATGATATT 195
Qy      158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      196 AGTGACTACCAAGAGATTATGATGATTTGGACGATGAGACTTTGATCGTTTATTA 255
Qy      178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSer 197

```

Db 256 AAAGGTGTTAGATAGAGTATGAAGCTTATTTAGATTGTTGTAATCATACATCT 315
QY AsnGluHisGlnTrp-----AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsn 215
Db 316 GATGAACATCTTGTTTATAGAAATCCAAATCTAGTAAAGACAATCCCAACGTCAT--- 372
QY PheTyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGlu 235
Db 373 ---TGGTACATTTGGCAAGATCCAAAG---CCAGAT----- 402
QY 236 IlePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGly----- 251
Db 403 -----GGCTCTGAACCTAACAACTGGGAAGTATATTAATGATCTACATGGGA 453
QY 252 -----ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeu 265
Db 454 TATGATGCTAATAGTACGAAATATTTCCATTTATTCAGTAAACAAACCTGATTTG 513
QY 266 AsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsn 285
Db 514 AATGGGGTAAACCGAAGTAGAGATGCTGTATTCGAAATGATGAACTGGTGGTTTGTAT 573
QY 286 LeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpIysGlnMetGlyThr 305
Db 574 AAAGGCATTTGATGGATTTAGAGTAGATGCAATTAACGCATATTAAGAAGACGTTTGAAGCG 633
QY 306 SerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArg 325
Db 634 GGTGACTTACCTGACTGAGGGTAAACATATATGCCAGCATTTGATGTAGTATGAAT 693
QY 326 -----IleAlaAlaProAla 330
Db 694 CAGCCAGGTATACAACTTGGTTACAAGAGATGAAGATCGCTCAATTAAGTAAATGATGAC 753
QY 331 ValPhePheIysSerGluAlaIle-----ValHisProAspGlnValValGlnTyrIleGly 349
Db 754 ATTATGACTGTGGTGAAGCGAATGTGTGAAGCCCTGATGATGCTGATGACTGGGTGGG 813
QY 350 GlnAspGluCysGlnIleGlyTyrAsnProLeu-----GlnMetAlaLeuLeuTrp 366
Db 814 GAAGAAATGGTAAA-----TTTATATCATATTCCAATTTGCAACATTTGGGACGTGG 867
QY 367 AsnThrLeuAlaThrArg---GluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHis 385
Db 868 AATAGTGGTGATTCCTCATTTTGTATGATAAATCGTATAAATCTGTATAATAGATGGCAA 927
QY 386 AsnLeuProGluHisThrAlaTrp---ValAsnTyrValArgSerHisAspAspIleGly 404
Db 928 AAACAACCTTGAATAAAGTTGGATGCGTTGTTTATTGAAATCATGACCAACCCGCA 987
QY 405 -----TrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHis 422
Db 988 CGTGTATCGACGTGGGTGACGATGCAAGTATTGG-----TATGAA--- 1029
QY 423 ArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyVal 442
Db 1029 ----- 1029
QY 443 ProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaAlaLeu 462
Db 1030 -----TCAGCAACAAGTCATCGAGCT 1050
QY 463 ValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleIysLeuLeuTyrSerIle 482
Db 1051 GTTATTTCTTGCACAA----- 1068
QY 483 AlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGly----- 499
Db 1069 -----GGTACGCATTCATTTATCAAGGTCAGAAATTTGATGACGAT 1113
QY 500 -----ThrLeuAsnAsp-----AspAspTrpSer--- 507
Db 507 ----- 507

Db 1114 TATCCATTTGAAGTATTGAACATTTAAACGATGCTGCTGTTAAAAATGACTATCAAT 1173
QY 508 -----GlnAspSerAsnIys 512
Db 1174 GTGAAAGCTCAAGCTGGAGATGTAGACGCTTTACTTTCGAAATATATAAAGATGAGAACCGA 1233
QY 513 SerAspAspSerArg-----TrpAlaHis----- 520
Db 1234 ---GATAATTTCTCGACACCAATGCAATGGATGATACGTTAAATGGAGGATTTACAAT 1290
QY 521 -----ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsn 533
Db 1291 GGTGAACCGTGTTCCTCCAGTGAATCCGAATTTATAAACTATCAATGTTGCTACATTA 1350
QY 534 AspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArg 553
Db 1351 GAAGATGACGATTCAGTATTACAAATTTTATAAAGATTTA-----ATTCAATTAAGA 1401
QY 554 GlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnIysHis 573
Db 1402 AAATCTAATGATGATATACGTATATGCTCAATTTGATTAGTAGATGCTGCTGAAATTCACAA 1461
QY 574 IleIleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSer 590
Db 1462 GTTTTTCATACATGAGAACATTAATAAGAAAGCAAGTTCTTATAGTAGGTAATCTTACT 1521
QY 591 GluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheIysAlaHisAsp 610
Db 1522 AACCAAGACGACGATTAACCT-----GTACCATTT-----GAT 1554
QY 611 LeuIleGlyGlyIysThrValSerLeuAsnGlnAspLeu-----ThrLeu 625
Db 1555 GTAAGTCATGAGAGAGTGAAGCTATTAATTTATGATGCCAAAGTTAATTTAAACAGTTG 1614
QY 626 GlnProTyrGln 629
Db 1615 GGTCCATATGAA 1626

RESULT 12
US-10-061-269-13
; Sequence 13, Application US/10061269
; Publication No. US20030087416A1
; GENERAL INFORMATION:
; APPLICANT: Mattes, Ralf
; Klein, Kathrin
; Schiweck, Hubert
; Kunz, Markwart
; Munir, Mohammed
; TITLE OF INVENTION: Preparation of Acariogenic Sugar
; Substitutes
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/061,269
; FILING DATE: 04-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/374,155
; FILING DATE: 22-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S

```

/ REGISTRATION NUMBER: 33,694
/ REFERENCE/DOCKET NUMBER: 05638.0006-00000
/ TELECOMMUNICATION INFORMATION:
/   TELEPHONE: (202) 408-4000
/   TELEFAX: (202) 408-4400
/ INFORMATION FOR SEQ ID NO: 13:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 1782 base pairs
/     TYPE: nucleic acid
/     STRANDEDNESS: single
/     TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/   NAME/KEY: misc RNA
/   LOCATION: 1237..1331
/   OTHER INFORMATION: /note= "N = Unknown"
/ SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-061-269-13

Alignment Scores:
Pred. No.:      5,35e-28      Length:      1782
Score:          313.50      Matches:      151
Percent Similarity: 35.34%      Conservative: 84
Best Local Similarity: 22.71%      Mismatches: 208
Query Match:      9.20%      Indels:      222
DB:                14      Gaps:        29

US-09-843-007a-2 (1-636) x US-10-061-269-13 (1-1782)

QY 77 TrpGlnSer-----TyrSerGlnArgAsnSerSerLeuLysAspIleAla 93
DB 100 TGGAAAGTCGTCTTCTATCAGGTCTATCCGCGCTCTTCAGGAT-----147
QY 94 ArgGluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyValCysTyrVal 113
DB 148 -----ACCAACGGTGATGGGATC-----165
QY 114 AspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeu 133
DB 166 -----GGCGATTTCAAAGACTGACGGAAGCTCGACTATCTCAAGGGGCTC 213
QY 134 GlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGly 153
DB 214 GGCATAGACGCCATCTGGATCAATCCACATTACGGCTCTCCC-----AACACCGATAT 267
QY 154 GlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeu 173
DB 268 GGCTACGATATCAGCGACTATCGAGAGGTTCATSAAGGAATATGGGACGATGGAGACTTC 327
QY 174 ArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePhe 193
DB 328 GATCGTCGTAGTGTGAGTGAAGAGCGCGGATCGCGCTCATGTGTGATGTCGTGATC 387
QY 194 AsnHisThrSerAsnGluHisGluTrp-----AlaGlnArgCysAlaAlaGlyAspPro 211
DB 388 AACCATTCGAGTGACCAACACGATGGTTCAGAGAGCAGCGCGGCTCCCAAGACAAATCCC 447
QY 212 LeuPheAspAsnPheTyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArg 231
DB 448 TACCGTGAC-----TATTATTCTGGCGTGACGGCAAA-----GACGGTCACGAG---492
QY 232 ThrLeuArgGluIlePheProAspGlnHisPro-----GlyGlyPheSerGlnLeu 248
DB 493 -----CCAAACAATTACCTTCTCTTCGCGGGTTCGGCATGGGAG 534
QY 249 GluAsp-----GlyArgTrpValTrpThrThrPheAsnSerPheGlnTrpAspLeu 265
DB 535 AAGGACCCCGTTAACCGGGCAATATTACCTGCGCAATATTTCGGTGTGTCAGCAGCCGATCTG 594
QY 266 AsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsn 285
DB 595 AACTGGGACACCCCGAGGCTTCGCGAGGAACTCTATGCGATCGTCGGTTCGGGTCGAC 654

```

```

QY 286 LeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLys-----301
DB 655 AAGGCGGTATCAGGCATCGGTTCTGATACGGTGTCTACTACTCGAAGACACCGGTTTC 714
QY 302 -----GlnMet-GlyThrSerCysGluAsnLeuProGlnAlaHisAl 315
DB 715 CCGGATCTGACACCGGAGCAGATGAAGAACTTCGCGGAGGCTATACCGAGCGCGAAC 774
QY 315 aLeuIle-----ArgAlaPheAsnAlaValMetArgIleAlaAl 328
DB 775 CTTTCATGTTACCTGCGAGAAATGACAGAGAGGTCTTCGATCATTTATGACGCGGTACG 834
QY 328 aproAlaValPhePheLysSerGluAlaIleValHisProAspGlnValGlnTyrIle 348
DB 835 GCCGGCG-----AAATCTTCGGCG-----CT 855
QY 348 eGlyGlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnTh 368
DB 856 CCGCTCAATCAAGTCCGCGCTCTTCATCGACACCGGAGGAAGAGCTGGATGCTTTC 915
QY 368 rLeuAlaThrArgGluValAsnLeuLeuHis-GlnAlaLeuThrTyrArgHisAsnLeuP 388
DB 916 ACCTTTCGATCTGATCCGTTATGATCGCGCACTGGATCGTTGGCATAC-----962
QY 388 roGluHisThrAlaTrpValAsnTyrValArgSerHisAsp-Asp-----402
DB 963 -----CATTCGCGTACTCTAGCGGACTTCGCTCAAGACGATCGATAAGTTCGACGCCATC 1017
QY 403 -----IleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGly 419
DB 1018 CGCGCGCATATGCTGG-----1035
QY 420 TyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAla 439
DB 1036 -----AACACGTCTCTCTCGGCAATCAGCAC-----1062
QY 440 ArgGlyValProPheGlnTyrAsnProSerThr-----GlyAspCysArgVal 455
DB 1063 -----AATCCCGCGGTATCGCATTTTGTGACGATCGCGCG 1101
QY 456 SerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArg 475
DB 1102 CAATGGCGCGAAGCTCG-----1119
QY 476 IleLysLeuLeuTyrSerIleAlaLeuSerThrGlyLeuProLeuIleTyrLeuGly 495
DB 1120 GCCAAGGCTGTCGCCCGCTCACCTTCACCCAGCAGGAGACCGCGTTCATCTTCCAAAGA 1179
QY 496 AspGluValGly-----ThrLeuAsnAspAspAsp-----505
DB 1180 GATGAATCGGAATGACCAACTACCCCTTCAAGACGCTGCGAGGACTTTGATGATATCANN 1239
QY 505 -----505
DB 1240 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1299
QY 505 -----505
DB 1300 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1359
QY 506 -----TrpSerGlnAspSerAsnLys-----SerAspAspSerArgTrp 518
DB 1360 CGCAGCGCTTTCATGGGATGACAGTGTCTAATCGGGGATTCAACAACCTGGCAAGCTTGG 1419
QY 519 AlaHis--ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArg---AsnAspProSer 536
DB 1420 CTAAGGTCAATCCAACTCACTGAGATCAACCGCGCGGGAATTTGGGATCTCTATAA 1479
QY 537 ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsn 556
DB 1480 TCG-----GTCTACAGCTTTTACCGCAACCTGATCTCAATCCGCGCATGAACCT 1527
QY 557 ProArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleLeuGly 576

```



```

1528 CCGCTCTTTCGACCGGAGCTATCCGACATCGATCCGAGTAATCCGATGCTCTATGCC 1587
QY
577 TyrIleArgAsn-----AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrPro 593
Db
1588 TATACCGCGACCGCAGGATCGCGAGCCTATCTGCTGTAGTCACTTCAAGCGAGGCCA 1647
QY
594 GlnThrValThr-----AlaHisThrLeuGlnAlaMetProPhe 606
Db
1648 AGAGATTTTCCAGCTTCCGAGCGCAGTCATATTCGCGAAACCTGATTGAGAGCAGTTCG 1707
QY
607 LysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsnGlnAspLeuThrLeuGln 626
Db
1708 CCAGCAGCTCCGCGCGCGCGGCTGCAAGC-----CTTGAGCTGCAG 1749
QY
627 ProTyrGln 629
Db
1750 CCTTGGCAG 1758

RESULT 13
US-10-081-872-191
; Sequence 191, Application US/10081872
; Publication No. US20030125534A1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Richardson, Toby
; APPLICANT: Frey, Gerhard
; APPLICANT: Short, Jay M.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Kerovuo, Janne S.
; APPLICANT: Slupaska, Malgorzata
; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: 09010-108001
; CURRENT APPLICATION NUMBER: US/10/081,872
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270,495
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/270,496
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/291,122
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 191
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Environmental
US-10-081-872-191

Alignment Scores:
Pred. No.: 1,83e-26 Length: 1596
Score: 300.50 Matches: 127
Percent Similarity: 36.73% Conservative: 75
Best Local Similarity: 23.09% Mismatches: 221
Query Match: 8.82% Indels: 127
DB: 14 Gaps: 20

US-09-843-007A-2 (1-636) x US-10-081-872-191 (1-1596)
QY
112 TyrValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLys 124
Db
169 TATGTAAATCTTTTACGATGCAATAAAGATGGACATGGTCACTTAAAGTCTTACA 228
QY
125 AspLysIleProTyrPhe-----GlnGluLeuGlyLeuThr 136
Db
229 CAAAGTTGCACTATTAAATGACGCAATCTTCATACAAAAAATGATCTTCAAGTAAC 288
QY
137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156
Db
289 GGGATTGATGATGTCAGTCACCCCTTCTCT-----AGCTATCATATAATATGAT 339
```

```

157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
Db
340 GTAAAGGACTATTAACAGTATCCGAGTAATCCGAGTAATCCGAGTAATCCGAGTAATCCGAGTA 399
QY
177 IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr 196
Db
400 ATGAAGAAGCAGACACAACAGAGACGTAAGCTTATATGACCTTGTGTGAATCATACG 459
QY
197 SerAsnGluHisGluThrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216
Db
460 AGCAGTGAACACCTTGGTTTCAAGTCGCTTAAAGATAAAACAGCAAGTACAGAT 519
QY
217 TyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236
Db
520 TACTATATTGGGCTGATAAA-----AATACCGACTTGAATGAA-- 558
QY
237 PheProAspGlnHisProGlyGlyPheSerGln-----LeuGluAspGly 251
Db
559 -----AAAGGATCTTGGGACACAACAGTATGCGCATAAAGCTCCAAACGGA 603
QY
252 ArgTrpValTyrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrp 271
Db
604 GAGTATTTTACCGAACGTTCTGGAGAGGATGCTGACTTAATATTAGGATAACCTGAA 663
QY
272 ValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeu 291
Db
664 GTAAAGAAAAGAAATCAITTAACGTCGAAAGTTTGGCTAAAGCAGGCGTTGACGGGTT 723
QY
292 ArgMetAspAlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuPro 311
Db
724 CGCTAGATGCTGCGCTTCATATTTTAAA-----GGTCAACAGCTGAAGCGCTAAG 777
QY
312 GlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaVal 331
Db
778 AAAAATATCTGTGTGGAATGAGTTAGAGATCGGATGAAAAAGAAAAATCCGAATGTA 837
QY
332 PhePheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGlnAsp 351
Db
838 TATCTAACGGGTGAAGTATGGATCAGCTGAGTGGTAGCTCCTTAT----- 885
QY
352 GlnCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThr 371
Db
885 ----- 885
QY
372 ArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391
Db
886 -----TATCAATCGCTTGATTCCTTTTATTAAATTTTGGTTTATGAGCAGGA 927
QY
392 AlaTrpValAsnTyrValArgSerHisAspAspIleGlyTyrThrPheAlaAspGluAsp 411
Db
928 AAAATTGTCACCTCTGTAAAGCAGGAATGATTCAGGA----- 966
QY
412 AlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheVal 431
Db
967 -----ATCGCCACTGTCAGCAGCAGCAACAGATGAACCTGTTCAAATCATCAATCCA 1017
QY
432 AsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGly 451
Db
1018 AACAAATTTGATGCAATTTCTTAACC-----AACCATGACCAAAATCGCGTC 1065
QY
452 AspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHis 471
Db
1066 ATGAGTCAGCTGAGCGGCGAT----- 1086
QY
472 AlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeu 491
Db
1087 ---GTGACAAAGCAAAATCAGCTGCTTCTATCTTACTTACCTTCCTCGCAACCCGTAT 1143
QY
492 IleTyrLeuGlyAspGluValGlyThrLeuAsnAsp-----AspAsp----- 505
Db
1144 ATTTATTACGCTGAAGAAATCGGCATGACCGCTGAAAAGCCTGATGAATTAATCCGTGAA 1203
QY
506 -----TrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHisArgPro 522
```

Db 1204 CCGTTCGGCTGG---TACGAAGGAAACGGACTTGGCAAAACCAGCTGG---GAAACACCT 1257
 Qy 523 ArgTyrAsn-----GluAlaLeuTyrAlaGlnArgAsnAsp 534
 Db 1258 GTATACAAATAAGCGGAAACGGTGTCTGTAGAGACACAACCAACCAAAAGGAT--- 1314
 Qy 535 ProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGln 554
 Db 1315 ---TCTTTGTTAAATCATTAACCGTGAAATGATTCGCGTGGCTCAG 1356
 Qy 555 SerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle 574
 Db 1357 CAGCATGAAGATGTAGTAAAGGACCTTCAATCTATTTTCAGTAGACAGTAAAGAGTT 1416
 Qy 575 IleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu 591
 Db 1417 GTTGCTTTATAGCGGTACGTATTAAGGCACTCCATTAGTGTGATCATATATTTCAAT 1476
 Qy 592 TyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeu 611
 Db 1477 CAACCGGTAAAGATATCTGTAGCAGCGAAAGGTAAATTTGTTGCTAGTGAAGAA--- 1533
 Qy 612 IleGlyGlyLysThrValSerLeuAsnGln 621
 Db 1534 ---GGTCTAGGAAGTCAAA---AATCAG 1557

RESULT 14

US-10-081-872-171
 ; Sequence 171, Application US/10081872
 ; Publication No. US20030125534A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Callen, Walter
 ; APPLICANT: Richardson, Toby
 ; APPLICANT: Frey, Gerhard
 ; APPLICANT: Short, Jay M.
 ; APPLICANT: Mathur, Eric J.
 ; APPLICANT: Gray, Kevin A.
 ; APPLICANT: Kerovuo, Janne S.
 ; APPLICANT: Slupska, Malgorzata
 ; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
 ; FILE REFERENCE: 09010-108001
 ; CURRENT APPLICATION NUMBER: US/10/081,872
 ; PRIOR FILING DATE: 2002-02-21
 ; PRIOR APPLICATION NUMBER: US 60/270,495
 ; PRIOR FILING DATE: 2001-02-21
 ; PRIOR APPLICATION NUMBER: US 60/270,496
 ; PRIOR FILING DATE: 2001-02-21
 ; PRIOR APPLICATION NUMBER: US 60/291,122
 ; PRIOR FILING DATE: 2001-05-14
 ; NUMBER OF SEQ ID NOS: 321
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 171
 ; LENGTH: 1431
 ; TYPE: DNA
 ; ORGANISM: Environmental
 US-10-081-872-171

Alignment Scores:
 Pred. No.: 1,78e-26 Length: 1431
 Score: 300.00 Matches: 122
 Percent Similarity: 36.88% Conservative: 72
 Best Local Similarity: 23.19% Mismatches: 208
 Query Match: 8.80% Indels: 124
 DB: 14 Gaps: 18

US-09-843-007a-2 (1-636) x US-10-081-872-171 (1-1431)

Qy 112 TyrValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLys 124
 Db 4 TATGTAACCTCTTTTACGATGCAATAAAGATGGACATGGTATTATAAAGGCTTACA 63

Qy 125 AspLysIleProTyrPhe-----GlnGluLeuGlyLeuThr 136
 Db 64 CAAAGCTTGGATTATTAAATGATGCAATTCCTCATCAAAAGATGATCTTCAAGTAAAC 123
 Qy 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156
 Db 124 GGATTTGGATGATGCGGTCACACCTTCTCC-----AGCTATCATAAATATGAT 174
 Qy 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
 Db 175 GTAACGGACTATTATATATTGATCCGAGTATGGAAATCTGCAAGATTTCGCAAACTG 234
 Qy 177 IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr 196
 Db 235 ATGAAGAAGACGATAAACGAGATGTAAAGTCAATTATGAGACCTCGTTGTGAATCATACG 294
 Qy 197 SerAsnGluHisGlnTyrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216
 Db 295 AGCGTGAACACCTTGGTTTCAAGCTGATTAAGAGATAAAACAGCAAGTACAGAGAT 354
 Qy 217 TyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236
 Db 355 TACTATATCTGGCTGTATAA-----AATACCGACTTGAATGAA--- 393
 Qy 237 PheProAspGlnHisProGlyPheSerGln-----LeuGluAspGly 251
 Db 394 -----AAGGATCTTGGGACAGCAAGTATGGCATTAAGCCCCAAACGGA 438
 Qy 252 ArgTyrValTyrThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsnProTyr 271
 Db 439 GAGTATTTTACGGAACGTTTGGGAAGGAATCCCGGACTTAAATTACGATAATCTCTGAA 498
 Qy 272 ValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAlaIleLeu 291
 Db 499 GTAAGAAAGAAATGATTAAACGTAGGAAGTTTGGCTAAAGCAAGGAGTTCAGCGGTT 558
 Qy 292 ArgMetAspAlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuPro 311
 Db 559 CGTCTAGATGCTCGCTTCATATTTTAA-----GGCCAAACACCTGAAGCGCTAAG 612
 Qy 312 GlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaProAlaVal 331
 Db 613 AAAAAATCTCTGTGTGGGAATGAAATTTAGAGATGCAATGAAAGAGGAAACCCCTACGTA 672
 Qy 332 PhePheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGlnAsp 351
 Db 673 TATCTAACGGGTGAAGTATGGATGCAACCGAGTAGTAGTCTCTTAC----- 720
 Qy 352 GluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThr 371
 Db 720 ----- 720
 Qy 372 ArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391
 Db 721 -----TATCATTCGCTGATCTTTTATTAACTTTGATTAGCAGGA 762
 Qy 392 AlaTyrValAsnTyrValArgSerHisAspIleGlyTyrThrPheAlaAspGluAsp 411
 Db 763 AAGATTGTAAACTCTGTAAAAATCAGGAATGATCAAGGA----- 801
 Qy 412 AlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheVal 431
 Db 802 -----ATCGGACTCGAGCGGCAACGGATGAACCTGTTCAAATCATACATCAATCCA 852
 Qy 432 AsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGly 451
 Db 853 AATAAAATTCACGGTATTTTCTTAACC-----AACCATGACCAAAATCGCGTC 900
 Qy 452 AspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHis 471
 Db 901 ATGAGTGAGCTAAGCGCGCAT----- 921
 Qy 472 AlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeu 491

922	--GTGAATAAAGCAAAGTCAGCTGCCTCTATCTTACTACGCTTCCTGGCAACCGGTAT	978
492	IlleTyLeuClgYAspGluValGlyThrLeuAsnAsp-----Aspasp-----	505
979	ATTATTACCGTGAAGAAATCCGATGACCCGGTGAAGAAAGCTGATGATTAATCCCGTGA	1038
506	-----TlpSerGlnAspSerAsnLysSerAspAspSerArgTTrpAlaHisArgPro	522
1039	CGGTTCCGCTGG-----TACGAAAGGAAACGGACTTCGACAAACCAGCTGG-----GAAACACCT	1092
523	ArgTyrAsn-----GluAlaLeuTyAlaGlnArgAsnAsp	534
1093	GTATACAACAAAGGCGGCAACGGCGTGTCTGTAGAAGCACAAACAAACAAAGGAC--	1149
535	ProSerThrAlaAlaGlyGlnIleTyGlnGlyLeuArgHisMetIleAlaValArgGln	554
1150	-----TCCTTGTAAATCATPACCGTGAATGATTCGCGTGGCGFCAG	1191
555	SerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle	574
1192	CAGCAGACAGATTAGTAAAGGAAACCGCTTCAATCTATTTCAGTAGACAGTAAGAAGTC	1251
575	IleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu	591
1252	GTTCCTTATAGCCGACGTATAAGGCCAAATCGATTAGCTGTATCATATATTTCAAAT	1311
592	TyrProGlnThrValThr	597
1312	CAACCCGTTAAAGTATCT	1329

RESULT 15

```

US-10-081-872-167
; Sequence 167, Application US/10081872
; Publication No. US20030125534A1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Richardson, Toby
; APPLICANT: Frey, Gerhard
; APPLICANT: Short, Jay M.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Kerouac, Janne S.
; APPLICANT: Slupska, Malgorzata
; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCES: 09010-108001
; CURRENT APPLICATION NUMBER: US/10/081,872
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270,495
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/270,496
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/291,122
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Environmental
US-10-081-872-167

```

Alignment Scores:	2.11e-26	Length:	1596
Pred. No.:	300.00	Matches:	122
Score:	3.88%	Conservative:	72
Percent Similarity:	2.13%	Mismatches:	208
Best Local Similarity:	8.80%	Indels:	124
Query Match:	14	Gaps:	18
DB:			

US-09-843-007A-2 (1-636) x US-10-081-872-167 (1-1596)

112 TyrValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLys 124
Qy :|||:::||||
Db TATGTAAACTCTTTTACGATGCAATAAAGATGGAGCATGGTGATTATAAAGCTCTTACA 228

125 AspLysIleProTyrPhe-----GlnGluLeuGlyLeuThr 136
Qy :|||:::||||
Db CAAAAGTTGGATTATTAAATGATGGCAATTCTCATACAAGAAGATGCATTCCTCAAGTAAC 288

137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156
Qy :::::|||:::
Db GCGAATTGGATGATGCCGTCAACCCTTCTCC-----AGCTATCATATAAATATGAT 339

157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
Qy :|||:::||||
Db GTAACGACGATTAATTAATATGATCGCGAGTAGGAATCTCGAAGATTTTCGCAAACTG 399

177 IleAlaAlaLeuHisGlnAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThr 196
Qy :|||:::||||
Db ATCAAGAAAGCACAGATAAACGAGATGTAAAGTCATTATGCACCTCGTTGTGAATCATACG 455

197 SerAsnGluHisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAsnPhe 216
Qy :|||:::||||
Db AGCAGTGAACACCTTGGTTTCAAGCTGCANTTAAACAGATAAAAAACGAAGTACAGAGAT 519

217 TyrTyrIlePheProAspArgAlaMetProAspGlnTyrAspArgThrLeuArgGluIle 236
Qy :|||:::||||
Db TACTATATCTGGCGCTGATAAA-----AATACCGACITTGAAATGAA-- 558

237 PheProAspGlnHisProGlyGlyPheSerGln-----LeuGluAspGly 251
Qy :|||:::||||
Db -----AAAGGATCTTGGGGACACCAAGATATGGCATAAAGCCCCCAAACGGA 603

252 ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTyr 271
Qy :|||:::||||
Db GAGTATTTTACGGACCTTTTGGGAAGGAATGCCGGACTTAATACGATAATCCTGAA 663

272 ValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeu 291
Qy :|||:::||||
Db GTAAGAAAGAAATGATTAACTAGGAAAGATTTTGGCTAAAGCAAGCAGTGTGACGGGTT 723

292 ArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsnLeuPro 311
Qy :|||:::||||
Db CGICTAGATCTGGCTTCATATTTTTTAAA-----GGCCAACACCTGAAGCGCTAAG 777

312 GlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaLaProAlaVal 331
Qy :|||:::||||
Db AAAAAATCTCCTGTGTGGAATGAATTTAGAGATGCAATGAAAAAGGAAAAACCCCTAACGTA 837

332 PhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGlyGlnAsp 351
Qy :|||:::||||
Db TATCTAAACGGGTGAAGTATGGATCAACCGGAGTAGTAGCTCCTTAC----- 885

352 GlucCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThr 371
Qy :|||:::||||
Db ----- 885

372 ArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391
Qy :|||:::||||
Db -----TATCAATCCCTTGATCTTTTATTAATCTTTGATTTTACGACGA 927

392 AlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPheAlaAspGluAsp 411
Qy :|||:::||||
Db AAGATTGTAACTCTGTAAATTCGGAATGATCAAGA----- 966

412 AlalaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheVal 431
Qy :|||:::||||
Db -----ATCGCGAGCTGCAGCAGCGCAACGGATGAATCTGTTCAAATCATACAATCCA 1017

432 AsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGly 451
Qy :|||:::||||
Db AATAAAATTTGCGGTATTTCTTAAACC-----AACCTAGCAAAATTCGGCTC 1055

452 AspCysArgValSerGlyThrAlaAlaAlaLeuValGlyLeuAlaGlnAspAspProHis 471
Qy :|||:::||||

```
Db      1066 ATGAGTGAGCTAAGCGCGAT----- 1086
Qy      472 AlaValAspArgIleLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeu 491
Db      1087 ---GTGATATAAGCAAGTACGCTGCCCTCTATCTTACTTACGCTTCTGGCAACCGGTAT 1143
Qy      492 IleTyrLeuGlyAspGluValGlyThrLeuAsnAsp-----AspAsp----- 505
Db      1144 ATTTATTACGGTGAAGAAATCGGCATGACCGGTGAAAAGCCGTGATGAGTTAATCCGTGAA 1203
Qy      506 -----TTPSerGlnAspSerAsnLysSerAspSerArgTTPAlaHisArgPro 522
Db      1204 CCGTTCGCTCG---TAGGAAGGAACGGACTTGGCAAAACCAGCTGG---GAAACACCT 1257
Qy      523 ArgTyrAsn-----GluAlaLeuTyrAlaGlnArgAsnAsp 534
Db      1258 GTATACAACAAAGCGGCAACGGCGTGTCTGTAGAGCACAAACAAACAAAGGAC--- 1314
Qy      535 ProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGln 554
Db      1315 -----TCTTTGTATAATCATTCGCGTGAATGATTCGCGTCGCTCAG 1356
Qy      555 SerAsnProArgPheAspGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle 574
Db      1357 CAGCACGAGAGTAGTAAAGGAAACGCTTCAATCTATTTCAGTAGACAGTAAAGAGTC 1416
Qy      575 IleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu 591
Db      1417 GTTCCTATAGCCGCGCAGCTATAAAGGCAAAATCGATTAGCGTGTATCATATAATTCAAT 1476
Qy      592 TyrProGlnThrValThr 597
Db      1477 CAACCGGTAAAGTATCT 1494
```

Search completed: November 9, 2003, 01:50:34
Job time : 6659 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - nucleic search, using frame_plus_p2n model
Run on: November 8, 2003, 19:48:18 ; Search time 2885 Seconds
(without alignments)
5357.937 Million cell updates/sec
Title: US-09-843-007A-2
Perfect score: 3408
Sequence: 1 MLPTQGVLIQLVLTIRL.....VSLNQDLTLPQVQWMLIEA 636
Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0
Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2 1/USPT0 spool p/US09843007/runat 07112003 170420 25360/app query.fasta_1.775
-DB=EST -OPMT=fastap -SUFFIX=p2n.rest -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd -LIST=45
-DOALIGN=200 -THRESHOLD=0.5 -THRESHOLD=0.5 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09843007/SCGN 1 1 3549 @runat 07112003 170420 25360 -NCPU=6 -ICPU=3
-NO MAP -LARGESQ -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -PGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_vrt:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	285.5	8.4	659	9	AI107032	AI107032 GH06318.5
2	278.5	8.2	641	9	AI292849	AI292849 GH15856.5
3	277	8.1	679	12	BM656725	BM656725 170006873
4	276.5	8.1	626	9	AI297229	AI297229 LP11488.5
5	274.5	8.1	638	9	AI402332	AI402332 GH21508.5
6	272.5	8.0	1981	28	BH770895	BH770895 LIMGtag62
7	271.5	8.0	673	12	BM610365	BM610365 170006871
8	271.5	8.0	712	12	BM646129	BM646129 170006873
9	265	7.8	785	29	BZ578463	BZ578463 msh2_5885
10	263	7.7	587	12	BM619567	BM619567 170006874
11	263	7.7	711	12	BM652118	BM652118 170006873
12	262.5	7.7	704	12	BM628529	BM628529 170006874
13	262.5	7.7	706	12	BM650445	BM650445 170006873
14	262.5	7.7	912	11	CNS08PQT	EX022785 Single re
15	260.5	7.6	694	12	BM579790	BM579790 170006872
16	257	7.5	611	12	BI507533	BI507533 BH170006874
17	256.5	7.5	648	12	BI583466	BI583466 RH22083.5
18	256.5	7.5	703	12	BM587500	BM587500 170006873
19	254.5	7.5	2271	11	AK052623	AK052623 Mus muscu
20	252.5	7.4	884	28	BH149112	BH149112 ENTQ104TF
21	252	7.4	718	12	BM555933	BM555933 170006872
22	251.5	7.4	684	12	BM635945	BM635945 170006871
23	251.5	7.4	698	12	BM616688	BM616688 170006873
24	251.5	7.4	711	12	BM641301	BM641301 170006873
25	250.5	7.4	604	9	AI238307	AI238307 GH14330.5
26	250.5	7.4	684	12	BM620803	BM620803 170006874
27	250	7.3	650	12	BJ329346	BJ329346 BJ329346
28	249.5	7.3	684	12	BM634536	BM634536 170006875
29	248.5	7.3	615	12	BI587736	BI587736 RH2884.5
30	248	7.3	688	12	BM624147	BM624147 170006874
31	248	7.3	692	12	BM637841	BM637841 170006875
32	248	7.3	693	12	BM604234	BM604234 170006870
33	248	7.3	693	12	BM622334	BM622334 170006874
34	248	7.3	694	12	BM597646	BM597646 170006875
35	246.5	7.2	621	12	BI634521	BI634521 RH3895.5
36	246.5	7.2	628	12	BJ365771	BJ365771 BJ365771
37	246	7.2	665	12	BM652834	BM652834 170006873
38	245.5	7.2	616	12	BI564072	BI564072 RH35755.5
39	245.5	7.2	623	9	AI389773	AI389773 GH21031.5
40	245.5	7.2	681	12	BM614246	BM614246 170006871
41	244	7.2	696	12	BM613854	BM613854 170006871
42	244	7.2	696	12	BM652741	BM652741 170006873
43	243	7.1	696	12	BM651858	BM651858 170006873
44	243	7.1	699	12	BM625632	BM625632 170006874
45	242.5	7.1	590	12	BJ324954	BJ324954 BJ324954

ALIGNMENTS

RESULT 1
AI107032
LOCUS GH06318.5prime GH Drosophila melanogaster head p0T2 Drosophila
DEFINITION melanogaster cDNA clone GH06318 5prime similar to V00204: Lvph
P890002570 PID:g7777 SWISS-PROT:P07190, mRNA sequence.
659 bp mRNA linear EST 19-APR-2001
ACCESSION AI107032
VERSION AI107032.1 GI:3475967
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 659)
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S., and Rubin, G.M.
BDGP/HEMI Drosophila EST Project
Unpublished
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>; est@fruitfly.berkeley.edu
hit genomic sequence DS04362; hit genomic sequence DS06189
Plate: 63 row: B column: 6
High quality sequence stop: 515.
Location/Qualifiers
1. .659
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GH06318"
/sex="male and female"
/dev_stage="adult"
/clone_lib="GH Drosophila melanogaster head pOT2"
/note="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2; Plasmid cDNA library."
BASE COUNT 148 a 160 c 191 g 159 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 6,16e-23 Length: 659
Score: 285.50 Matches: 65
Percent Similarity: 54.75% Conservativeness: 33
Best Local Similarity: 36.31% Mismatches: 76
Query Match: 8.38% Indels: 5
DB: 9 Gaps: 3
US-09-843-007A-2 (1-636) x AJ107032 (1-659)
QY 118 GlyAspLeuHisGluAlaValSerAspGlyLeuThrTyr 137
DB 129 GCGGATTTGAAGGAAATCCCTCCAGCTGCCTATCTGGCGGACACTGGGATCAGGCC 188
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157
DB 189 ACCTGGTTGAGTCCCATTTTTCAGTCGCC-----ATGATTGACTTTGGCTATGATATA 242
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValle 177
DB 243 TCGGATTACAGGCAATCCAGCGGAGTATGCGACCATGCGAGGATTCGAGGAGCTGATC 302
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
DB 303 GACACGGCTTCGAGCTGGGATCAGGTGTTCTGGACTTTGTCCGGAATCAGCTCG 362
QY 198 AsnGluHisGluThrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 217
DB 363 GATCAGCATGAGTGGTTCAAGAGTCTGCGCAAGGAGCGGGTTACGAGGATTC--- 419
QY 218 TyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhe 237
DB 420 TATGTGTGCACATGCGATGTCGAGGAGATGCGACTCGGTCCACCCACACACTGG 479
QY 238 ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg-----TrpValTrp 255
DB 480 CCATCGGTGTTCTACGGATCGCTTGGGAGTGGCAGAGGTCTGTCAGCAGTATTACCTG 539
QY 256 ThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAla 275
DB 540 CACCAGTTCACCAAGAGCAGCGGACATTGAATATGCGATATGCCAAAGTGGTTTCAGGCC 599

QY 276 MetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp 294
DB 600 ATGATGATGATGTTGCTTTTCTGGCTCACAAGGCGGTANCTGGTTTTCGATCATCAT 656
RESULT 2
AI292849
LOCUS
DEFINITION
AI292849 641 bp mRNA linear EST 19-APR-2001
GHI5856.Sprime GH Drosophila melanogaster head pOT2 Drosophila
melanogaster cDNA clone GHI5856 Spprime similar to V00204: LvpH
F8gn0002570 PID:57777 SWISS-PROT:P07190, mRNA sequence.
AI292849
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 641)
AUTHORS
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S., and Rubin, G.M.
BDGP/HEMI Drosophila EST Project
Unpublished
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>; est@fruitfly.berkeley.edu
Plate: 158 row: E column: 8
High quality sequence stop: 477.
Location/Qualifiers
1. .641
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GHI5856"
/sex="male and female"
/dev_stage="adult"
/clone_lib="GH Drosophila melanogaster head pOT2"
/note="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2; Plasmid cDNA library."
BASE COUNT 147 a 157 c 186 g 151 t
ORIGIN
Alignment Scores:
Pred. No.: 3.99e-22 Length: 641
Score: 278.50 Matches: 63
Percent Similarity: 54.91% Conservativeness: 32
Best Local Similarity: 36.42% Mismatches: 73
Query Match: 8.17% Indels: 5
DB: 9 Gaps: 3
US-09-843-007A-2 (1-636) x AI292849 (1-641)

QY 118 GlyAspLeuHisGluAlaValSerAspGlyLeuThrTyr 137
DB 130 GCGGATTTGAAGGAAATCCCTCCAGCTGCCTATCTGGCGGACACTGGGATCAGGCC 189
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157
DB 190 ACCTGGTTGAGTCCCATTTTTCAGTCGCC-----ATGATTGACTTTGGCTATGATATA 243
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValle 177
DB 244 TCGGATTACAGGCAATCCAGCGGAGTATGCGACCATGCGAGGATTCGAGGAGCTGATC 303
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
DB 304 GACACGGCTTCGAGCTGGGATCAGGTGTTCTGGACTTTGTCCGGAATCAGCTCG 363

FEATURES
source

Location/Qualifiers
1..626
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="Df11488"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
POT2"
/note="Organ: whole body; Vector: POT2; Site: 1: EcoRI;
Site: 2: XhoI; Sized fractionated cDNAs were directly
ligated into POT2. Plasmid cDNA library."

BASE COUNT
142 a 153 c 181 g 150 t

ORIGIN

Alignment Scores:
Pred. No.: 6,65e-22 Length: 626
Score: 276.50 Matches: 62
Percent Similarity: 55.29% Conservative: 32
Best Local Similarity: 36.47% Mismatches: 71
Query Match: 8.11% Indels: 5
DB: 9 Gaps: 3

US-09-843-007A-2 (1-636) x AI297229 (1-626)

QY 118 GlyAspLeuGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
Db 124 GCGATTTGAAAGGAATCACTCCAGCTCGGTATCTGGCGACACTGGCATCGCGCC 183
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
Db 184 ACCTGTTGAGTCCCATTTTTCAGTCGCCC-----ATGATTGACTTTTGGCTATGATATA 237
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 238 TCGATTACAGGCATCCAGCGGAGTATGCGACCATGCGAGTTTCAGGAGCTGATC 297
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSer 197
Db 298 GACACGGCTTTCAGCTGGGCAATCAAGTTGTTCTGGACTTTTGGCGAATCACAGCTCG 357
QY 198 AsnGluHisGluThrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 217
Db 358 GATCAGCATGAGTGTTCAGAGAGTCTGGCGAAGGAGCGCGGTTACAGAGTTTC--- 414
QY 218 TyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhe 237
Db 415 TATGTCTGGCAGCATGCGATCGTCCAGGAGATGGCACTCGGTGGCCACCCCAACTGG 474
QY 238 ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg-----TrpValTrp 255
Db 475 CCATCGGTGTTCAGCATCGCTTGGAGTGGCAGAGAGTCTGAGCAGTATACCTG 534
QY 256 ThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAla 275
Db 535 CACCAGTTTCCACCAAGGAGCAGCGGACTTGAATATCGCAATCCCAAGTGGTTACGGCC 594
QY 276 MetAlaGlyGluMetLeuPheLeuAlaAsn 285
Db 595 ATGGATGATGTGTGTCTTTCTTCTGCTCAAC 624

RESULT 5
AI402332
LOCUS
DEFINITION
GH21508.5prime GH Drosophila melanogaster head POT2 Drosophila
melanogaster cDNA clone GH21508 5prime similar to V00204: LVPH
F8gn0002570 PID:g77777 SWISS-PROT:P07190, mRNA sequence.
AI402332
AI402332.1 GI:4245419
EST.

SOURCE
ORGANISM

Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 638)
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic sequence AC06240
Plate: 215 row: A column: 8
High quality sequence stop: 551.

FEATURES
source

1..638
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GH21508"
/sex="male and female"
/dev_stage="adult"
/lab_host="DHS - alpha"
/clone_lib="GH Drosophila melanogaster head POT2"
/note="Organ: head; Vector: POT2; Site: 1: EcoRI; Site: 2:
XhoI; Sized fractionated cDNAs were directly ligated into
POT2. Plasmid cDNA library."
BASE COUNT 145 a 156 c 186 g 151 t
ORIGIN

Alignment Scores:
Pred. No.: 1.18e-21 Length: 638
Score: 274.50 Matches: 62
Percent Similarity: 54.65% Conservative: 32
Best Local Similarity: 36.05% Mismatches: 73
Query Match: 8.05% Indels: 5
DB: 9 Gaps: 3

US-09-843-007A-2 (1-636) x AI402332 (1-638)

QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
Db 130 GCGATTTGAAAGGAATCACTCCAGCTCGGTATCTGGCGACACTGGCATCGCGCC 189
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
Db 190 ACCTGTTGAGTCCCATTTTTCAGTCGCCC-----ATGATTGACTTTTGGCTATGATATA 243
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 244 TCGATTACAGGCATCCAGCGGAGTATGCGACCATGCGAGTTCGAGGAGCTGATC 303
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSer 197
Db 304 GACACGGCTTTCAGCTGGGCAATCAAGTTGTTCTGGACTTTTGGCGAATCACAGCTCG 363
QY 198 AsnGluHisGluThrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 217
Db 364 GATCAGCATGAGTGTTCAGAGAGTCTGGCGAAGGAGCGCGGTTACAGGATTC--- 420
QY 218 TyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhe 237
Db 421 TATGTCTGGCAGCATGCGATCGTCCAGGAGATGGCACTCGGTGGCCACCCCAACTGG 480
QY 238 ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg-----TrpValTrp 255
Db 481 CCATCGGTGTTCAGCATCGCTTGGAGTGGCAGAGAGTCTGAGCAGTATACCTG 540

QY 256 ThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAla 275
 DB 541 CACCAGTTCCACGAGCAGCGGACTTGAACATATCGCAATCCCAAGTGTTCAGGCC 600
 QY 276 MetAlaGlyGluMetLeuPheLeuAlaLeuGly 287
 DB 601 ATGGATGATGTTGCTTTCTGGCTCAACAGGCG 636

RESULT 6
 BH770895/c
 LOCUS BH770895
 DEFINITION LLMGT:ag629 MG1363 Random Sequence Tag Library Lactococcus lactis
 subsp. cremoris genomic, genomic survey sequence.
 ACCESSION BH770895
 VERSION BH770895.1
 KEYWORDS GI:20373852
 SOURCE Lactococcus lactis subsp. cremoris
 ORGANISM Lactococcus lactis subsp. cremoris
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Lactococcus.

REFERENCE
 1 (bases 1 to 1981)
 AUTHORS Bolotin,A., Ehrlich,S.D. and Sorokin,A.
 TITLE Studies of genomes of dairy bacteria Lactococcus lactis
 JOURNAL Sci. Aliments, (2002) In press
 COMMENT Contact: Sorokin A
 Genetique Microbiome
 INRA
 CRG INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
 Tel: 33 1 34 65 25 16
 Fax: 33 1 34 65 25 21
 Email: sorokine@jouy.inra.fr
 best homologue in strain IL1403 is amyY (90%)
 Class: shotgun
 High quality sequence start: 30
 High quality sequence stop: 1953.
 Location/Qualifiers
 1..1981
 /organism="Lactococcus lactis subsp. cremoris"
 /mol_type="genomic DNA"
 /strain="MG1363"
 /db_xref="taxon:1359"
 /clone_lib="MG1363 Random Sequence Tag Library"
 /note="Vector: pSGM2; Site 1: Smar; Library of
 chromosomal fragments of L.lactis strain MG1363 was
 prepared by partial AluI digestion or by sonication."
 BASE COUNT 574 a 429 c 335 g 643 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.22e-20 Length: 1981
 Score: 272.50 Matches: 136
 Percent Similarity: 38.42% Conservative: 93
 Best Local Similarity: 22.82% Mismatches: 241
 Query Match: 8.00% Indels: 127
 DB: 28 Gaps: 22

US-09-843-007A-2 (1-636) x BH770895 (1-1981)

QY 75 GlnAlaTrpGlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAlaArg 94
 DB 1885 AGGCTTCTCAAAAGTAACTTCAAAAGATTCACGATGAAAAGCGGCTGTCAGTCAA 1826
 QY 95 GluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyValCysTyrValAsp 114
 DB 1825 AAGGTTGATACGAGTTTGTATCGTAATTTTATGAAATA-----TTTACGAGT 1778
 QY 115 LeuPheAla-----GlyAspLeuLysGlyLeuLysAspLysIle 127
 DB 1777 TCATTGGCGATTCAAAATCATGATGTTGAGGCGGATTAAATGGGTCACGCAACATTTA 1718
 QY 128 ProTyrPheGln-----GluLeuGlyLeuThrTyrLeuHis 139
 DB 1717 GATTATCTCAATACGGGGAATAAAATTTCTCAACAGATTTGAAAGTCCAAAGGACTTTGG 1658

QY 140 LeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSer 159
 DB 1657 ATGACCCCAATCTTTGCTAGTCCA-----AGCTATCATGCTATGATGTCCAGCAAT 1607
 QY 160 TyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAla 179
 DB 1606 TACGAAGAAATTAATCTCTAAATTTGGCACAATGCTGATTTGAAAATCTGATGCTCTCAG 1547
 QY 180 LeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSerAsnGlu 199
 DB 1546 GCTAAATAACGAGGAATTTGCGTGATTTGGACATGCGGTTATTCACACGCGACTGAC 1487
 QY 200 HisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIle 219
 DB 1486 AATCTCTGGTTTCAAAAAGCAGCTAGCTCGCACAAGAAATATATATGCGTTATTAT- 1433
 QY 220 PheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAsp 239
 DB 1432 -----AATTGTCGACACAGCAAGAG----- 1409
 QY 240 GlnHisProGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrThrPheAsn 259
 DB 1408 -----GGTTATAGCTTAGCCAGCAACGCAATACTATGAATCAGAGTTTCAT 1361
 QY 260 SerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGlu 279
 DB 1360 AAATCGATGCCAGCACCTCAATCTTGGCAATCTCGAGTCAAAAAGAAATTTGAAAATT 1301
 QY 280 MetLeuPheLeuAlaAsnLeuGlyValAlaPheIleLeuArgMetAspAlaVal-AlaPheIle 299
 DB 1300 ACGAAATTTGGTTAGACAAAAGGTGTTCTCGTTTACGTTTGGAGCGTGTGGTTTTTAT 1241
 QY 299 eTrpLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAl 319
 DB 1240 TTCTCAAAATGATGATAAAAAACGACGCGCTTTACCAAAATGGCTGCAGC-----ATACGCTC 1184
 QY 319 aPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleVal 339
 DB 1183 AAAAGCAAAATCCAAAAGCTTATTAGTGGTGAAGTTTTTTTCAA----- 1138
 QY 339 LHisProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnPr 359
 DB 1137 -CGGCTTCGGCAATGTATGATATTATTCAGTCAG-----GTGTTGATCTTTATTTTCAT 1085
 QY 359 oLeuGlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuHisGlu 379
 DB 1084 TTTCRAATG-----CTTTGCAATCGAGTTCGAGTCCGATTATTCGGCTA----- 1039
 QY 379 nAlaLeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSe 399
 DB 1039 ----- 1039
 QY 399 rHisAspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrIleuGlyIleSerGl 419
 DB 1038 -----TGGCTTTGGCGAAGGCTCATTTGTTGGTCTCAG-CAAAATCGAAGC 996
 QY 419 yTyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAl 439
 DB 995 TTGGGACCAAGAA-----ATTCATGCTGCACACCGCAAGC 960
 QY 439 aArgGlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAl 459
 DB 959 TATTGACGCTCCCTTTTGTAGTAATCATGCACA-----GACCGGTGAGC 915
 QY 459 aAlaAlaValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLe 479
 DB 914 TAGCTTTTAAGCTCTTTAGCCATGCAA-----AAAATGGC 879
 QY 479 uTyrSerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGI 499
 DB 878 TGCTCAAACTTACTCTTTTGATGCGCGGAATCCATTATTATTACTACGCGGAAGATTGGG 819

QY 499 yThrLeuAsnAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyr-- 518
Db 818 GCTGACAGGTTCT-----GGTATTGACCAACAAACGCTTCCCAATCCCATGGTC 768
QY 519 -----AlaHisArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspPr 535
Db 767 GCGTATGATGAGCAGATTCCTCCCTAAGCAAAATGTCTGCTCCGGTGGAGCAGTAAACGGC 708
QY 535 oSerThrAlaAlaGlyGlnIleTyrGln-----G1 545
Db 707 GACNACTGACGAGGATCGGTGGTCAGCAAGAACGAAATCCGAATTCGCTACTCAATTG 648
QY 545 yLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyArgLeu-- 564
Db 647 GTACAAAATAATCTTCGGGTCAAGCGAAATATCTGAATTTGGCACTGCCAGGCTGAA 588
QY 565 ----ValThrPheAsnThrAsnLysHisIleIleGlyTyrIleArgAsnAlaLe 583
Db 587 AAATATTTCTGACGTAAGTAGCTTGTCTGTCTATTAATTAATGGAAGGAT-----TT 534
QY 583 uLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAl 603
Db 533 GACCAATTGCAATAATTTTTCAGAT---AAAGAAAGTATGACAGTGAAACTCCCGAAGG 477
QY 603 aMetProPheLys-----AlaHisAspLeuIleGlyLysThrValSerLeuAsnGl 621
Db 476 AGTTGTGGCAAAATAATTACTGACCAACTTTCTGGTGGTGGAGTCCAAAGCGACATTGAC 417
QY 621 nasp-----LeuThrLeuGlnProTyrGlnValMetTyrLeuGlu 634
Db 416 TGAATGGAATAATGACAAATTCAGCTTACAGTACAGTATTTTTGAAA 371

RESULT 7
BM610365
LOCUS
DEFINITION
17000687110664 A.Gam.ad.cdna.blood1 Anopheles gambiae cdna clone
19600449688657 5', mRNA sequence.

ACCESSION
BM610365
VERSION
BM610365.1 GI:18908469
KEYWORDS
EST.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicidae;
Anopheles.
REFERENCE
1 (bases 1 to 673)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charl-
lab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AY2 row: p column: 19
Seq primer: M13 Reverse.

FEATURES
source
1..673
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449688657"
/dev_stage="Adult"
/lab_host="DHI0P"
/clone_lib="A.Gam.ad.cdna.blood1"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.

Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT 154 a 174 c 197 g 148 t
ORIGIN
Alignment Scores: 2.92e-21 Length: 673
Score: 271.50 Matches: 65
Percent Similarity: 47.87% Conservative: 36
Best Local Similarity: 30.81% Mismatches: 69
Query Match: 7.97% Indels: 41
DB: 12 Gaps: 7

US-09-843-007A-2 (1-636) x BM610365 (1-673)
QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPhe---GingluLeuGlyLeuThr 136
Db 38 GGAGATTTCGCGCGCAATAATGAAAAAGTTCCGTACCTGCGACGAGCTCGCATCGAT 97
QY 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAla 156
Db 98 GCATATGGCTGTGCGCGCATCTTCAAGTCACCG-----ATGGCGGACCTTTGGCTAGCAT 151
QY 157 ValSerGlyTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
Db 152 ATTCGCGATTTCGCGGACATTCATTCGGAGTTTGGACAAATGCTGACCTGGAGAGCTG 211
QY 177 IleAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThr 196
Db 212 CGGACCGCTTCTAATTCAGAGGGGCTCAAGCTAATCTCTGGACTTTGTTCCGAACACAGC 271
QY 197 SerAsnGluHisGluTyrAlaGlnArgCysAlaAlaGlyAspProLeuPheAsnAspPhe 216
Db 272 AGCCATGAATGAGTGGTGTCTTAAAGTCCCTCCAGAGGATCCCACTACAGCATATAC 331
QY 217 TyrTyrIlePheProAspArg-----ArgMetProAspGlnTyrAsp 230
Db 332 TATGTGTGCATCCCGGTAAAGACGCTGGCTAAACGGAGCGCGCTGCCT----- 379
QY 231 ArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe---SerGlnLeuGlu 249
Db 380 -----CCTTCGAACTGGTGAGTGTTCTTCCTAGTT 406
QY 250 AspGlyArgTyrValTyrThr-----ThrPheAsnSer 260
Db 407 GGTTCAGCTCGGAGTGGACGACGCTGGAGGAGTACTATCTCCACCACTTCTAGTT 466
QY 261 PheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet 280
Db 467 AAGCAGCCCGATCTCACTATCGCAATCCCGCGCTAGTACAGAAATGAAAGACGTGATG 526
QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr 300
Db 527 AGTTCTGGCTTGGCAAGGCGTCAATGAGTTCGATCGATCGCTGCGCTGCTATCTTTT 586
QY 301 LysGlnMetGlyThrSerCysGluAsnLeuPro 311
Db 587 -----GAATCGCTGCC 598

RESULT 8
BM646129
LOCUS
DEFINITION
17000687319533 A.Gam.ad.cdna1 Anopheles gambiae cdna clone
19600449637968 5', mRNA sequence.
ACCESSION
BM646129
VERSION
BM646129.1 GI:18945640
KEYWORDS
EST.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 712)

AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab
, R., Collins, P.H., Venter, J.C. and Hoffman, S.L.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AYS row: A column: 06
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source
1..712
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449637968"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="A.Gam.ad.cdna1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.wf4.org)."
BASE COUNT 156 a 181 c 201 g 174 t
ORIGIN

Alignment Scores:
Pred. No.: 3,19e-21 Length: 712
Score: 271.50 Matches: 65
Percent Similarity: 47.87% Conservative: 36
Best Local Similarity: 30.81% Mismatches: 69
Query Match: 7.97% Indels: 41
DB: 12 Gaps: 7

US-09-843-007A-2 (1-636) x BM646129 (1-712)

Qy 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPhe---GlnGluLeuGlyLeuThr 136
Db 137 GGAGATTTCGGCGGATTAATGAAAAGTTCCGTACCTCGCAGCGAGCTCGGCATCGAT 196
Qy 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAla 156
Db 137 GCATATGCTGCTGCCGATCTTCAAGTCACG-----ATGGCGGACTTTGGCTACGAT 250
Qy 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
Db 251 ATTGCGGATTCGGCGACATTCATTCGGAGTTTGGAACAATTCGTACCTGGAAGCGCTG 310
Qy 177 IleAlaAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThr 196
Db 311 CGGACCGTTGTAATGACAGAGGGCTCAAGCTAATCCTGGACTTTGTTCCGAACACACG 370
Qy 197 SerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216
Db 371 AGCGATGAAGTGAAGTGGTTCTTAAAGTCGTCACAGAGGATCCACCTACAGCGATTAC 430
Qy 217 TyrTyrIlePheProAspArg-----ArgMetProAspGlnTyrAsp 230
Db 431 TATGTGTGGCCACCGCGGTAAGACGCTGCTAAGCGGACGCGGTCCT-----478
Qy 231 ArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe---SerGlnLeuGlu 249
Db 479 -----CCTTCGAAGTGGGTGAGTCTTCGTT 505
Qy 250 AspGlyArgGtrValTrpThr-----ThrPheAsnSer 260
Db 506 GGTTCAGCTGCGGAGTGAAGACGACGTCGGGAAGGAGTACTATCTCCACGATTCCTGGTT 565

Qy 261 PheGlnTrpAspLeuAsnTyrSerAsnProTyrPValPheArgAlaMetAlaGlyGluMet 280
Db 566 AAGCAGCCGATCTGAACATATCCGCGCTAGTACAGAAATCAAGATCTGA 625
Qy 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleP 300
Db 626 ACGTTCGTGGCTTGCAAGAGGCTCATGGATTCGCATCGCGTCCGTAICTTTT 685
Qy 301 LysGlnMetGlyThrSerCysGluAsnLeuPro 311
Db 686 -----GAATCGTGCCT 697

RESULT 9
BZ578463 785 bp DNA linear GSS 17-DEC-2002
LOCUS msh2.5885.x1 msh Pseudomonas aeruginosa genomic clone msh2.5885,
DEFINITION genomic survey sequence.

ACCESSION BZ578463
VERSION BZ578463.1 GI:27213524
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 785)

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

CONTACT: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062218554

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES Location/Qualifiers
source
1..785

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="MSH"

/db_xref="taxon:287"

/clone="msh2.5885"

/clone_1lb="msh"

/notes="Environmental isolate. Whole genomic shotgun

library."

BASE COUNT 154 a 274 c 224 g 131 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 2.2e-20 Length: 785

Score: 265.00 Matches: 74

Percent Similarity: 45.78% Conservative: 40

Best Local Similarity: 29.72% Mismatches: 96

Query Match: 7.78% Indels: 40

DB: 29 Gaps: 10

US-09-843-007A-2 (1-636) x BZ578463 (1-785)

Qy 174 ArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePhe 193

Db 90 CGCGGTTCATCCCGAGGCGCCATCGACGGTCTCGGGGGATTACCGAGCTGGTATC 149

Qy 194 AsnHisThrSerAsnGluHisGluTrp-----AlaGlnArgCysAlaAlaGlyAsp 210

Db 150 ACCACACTCCGACGACGATCCCTGGTTCATCCGCGCCCGCCAGGAGGATCG 209

Qy 211 ProLeuPheAspAsnPheTyrIlePheProAspArgMetProAspGlnTyrAsp 230

Db 210 CGCGCCCGCGAC-----TACTAGTCTGTCGGAC-----ACCGACGAGAAATACCG 257

```

QY 231 ArgThrLeuArgLulePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAsp 250
DB 258 GGTACG---CGGATCATCTTCATCGAC-----ACCGAGCAG 290
QY 251 GlyArgTrpValTrpThr-----ThrPheAsnSerPhe 261
DB 291 TCAACTGGACCTGGACCGGTGGCCACAGACTACTTGGACCGCTTCTATCCAC 350
QY 262 GlnTrpAspLeuAsnTrpSerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeu 281
DB 351 CAGCGCGACCTGAACCTCGACACACCGGAGTCTCGCGAGGTGCTGGGGGTGATGGC 410
QY 282 PheLeuAlaAsnLeuGlyValAspLeuLeuArgMetAspAlaValAlaPheLeuTrpLys 301
DB 411 TACTGGCTGGACATGGCGTGGACGCTCGCGCTGGACCGGATCTCTACCTGATCGAA 470
QY 302 GlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuLeuArgAlaPheAsn 321
DB 471 CGCGACGGACACGAGCAGGAGAACCTGCGGAGACCCACGAGTCTCAAGCGCATCGC 530
QY 322 AlaValMetArgLeuAlaAlaProAlaValPhePheLysSerGluAlaLeuValHisPro 341
DB 531 GCGGAGCTGGACGGCGACTATCCGACCGCATCTGCTGCGCGAGCGCCACCGATGGCG 590
QY 342 AspGlnValValGlnTrpIleGly-----GlnAspGluCysGlnIle 355
DB 591 GAGACACCGCGCGGACTCTCGCTGCGAGGATGGCGGTAGCGGAGCGAGATGCCATG 650
QY 356 GlyTyrAsn---ProLeuGlnMetAlaLeuLeuTrp-----AsnThrLeuAlaThr 371
DB 651 GCCTTTCACCTTCGCTGATCGCGCATGTATCATGTGCGCATCCGCCAGGAGATCGCTATC 710
QY 372 ArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391
DB 711 CGATC-ACCGACATCTTGGCGCCAGACCTC-----GGCATTCGCTCCATTCG 757
QY 392 AlaTrpValAsnTrpValArgSerHis 400
DB 758 CAATGGCGGATCTTCTGCGCAACAC 784

RESULT 10
BM619567
LOCUS 587 bp mRNA linear EST 25-FEB-2002
DEFINITION 17000687375124 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
19600449632028 5', mRNA sequence.
ACCESSION BM619567
VERSION BM619567.1 GI:189517985
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM
REFERENCE 1 (bases 1 to 587)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.D.
TITLE Celexa Anopheles gambiae EST project
JOURNAL Unpublished
COMMENT Contact: Holt R.A.
Celexa Genomics
45 W. Gude Dr.
Rockville, MD 20850, USA
Tel: 2404531151
Fax: 2404534580
Email: HoltRA@celexa.com
Plate: NU01004N8D row: N column: 14
Seq primer: M13 Reverse.
Location/Qualifiers
FEATURES
source
1..587
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"

```

```

/db_xref="taxon:7165"
/clone="19600449622928"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="A.Gam.ad.cDNA1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.m4.org)."
BASE COUNT 135 a 151 c 168 g 133 t
ORIGIN
Alignment Scores: 2,418-20 Length: 587
Pred. No.: 263.00 Matches: 62
Score: 47.96% Conservative: 32
Percent Similarity: 31.63% Mismatches: 68
Best Local Similarity: 7.72% Indels: 34
Query Match: 12 Gaps: 6
DB:
US-09-843-007a-2 (1-636) x BM619567 (1-587)
QY 118 GlyAspLeuLysGlyLeuLysAspLysileProTyrPhe---GlnGluLeuGlyLeuThr 136
DB 44 GGAGATTTCCGCGGCATATATGGAAGTTCCGTACCTCCGACGGAGCTCGGCATCGAT 103
QY 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156
DB 104 GCATATGGCTGTCGCGCATCTTCAAGTCACCG-----ATGCCGACCTTGGTACGAT 157
QY 157 ValSerSerTyrAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
DB 158 ATTCGCGATTTCCGCGACATTCATTCGGAGTTTGGACAAATTCGTGACCTGGAAGCGCTG 217
QY 177 IleAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThr 196
DB 218 GGCACCGCTTGTATGACAGAGCGGCTCAAGCTTAATCTGGACTTTGTTCCGAACACACG 277
QY 197 SerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsn 216
DB 278 AGCGATGAAGTGGTGGTTTCTTAAGTCCGTCAGAGAGGATCCACCTACAGCGATTAC 337
QY 217 TyrTyrIlePheProAspArg-----ArgMetProAspGly TyrAsp 230
DB 338 TATGTGTGGCATCCGCGTAAGACGCTGGCTAACGGGACGCGCGTGCCT-----385
QY 231 ArgThrLeuArgLulePheProAspGlnHisProGlyGlyPhe---SerGlnLeuGlu 249
DB 386 -----CCTTCGACTGGTGGTGTCTCCCGT 412
QY 250 AspGlyArgTrpValTrpThr-----ThrPheAsnSer 260
DB 413 GGTTCAGCTGGGAGTGGAAACGACGTGGGAGGAGTACTATCTCCACCGATTCTAGTT 472
QY 261 PheGlnTrpAspLeuAsnTrpSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280
DB 473 AAGCAGCCCGATCTGAACATATCGCAATCCGCGCTAGTACAGAAATGAAGACGTTGATG 532
QY 281 LeuPheLeuAlaAsnLeuGlyValAspLeuLeuArgMetAspAlaVal 296
DB 533 ACCTTCTGGCTGGCAAGGCGTCCATGATTCGCGATCGATCGCGTG 580

RESULT 11
BM652118
LOCUS 711 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687375124 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
19600449632068 5', mRNA sequence.
ACCESSION BM652118
VERSION BM652118.1 GI:18951629
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)

```


ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Neophlebotominae

REFERENCE 1 (bases 1 to 711)
AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab,
R., Collins, P.H., Venter, J.C. and Hoffman, S.B.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltR@celera.com
Plate: NU010049VH row: K column: 10
Seq primer: M13 Reverse.

FEATURES
source
1..711
 /organism="Anopheles gambiae"
 /mol_type="cDNA"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="1960049632068"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /clone_lib="A.Gam.ad.cdNAL"
 /note="Vector: pSPort1; Site 1: SalI; Site 2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen.
 cDNA inserts >500 bp cloned directionally into pSPort 1.
 Not 1 site is 3'. Clones available through the Malaria
 Research and Reference Reagent Resource Center
 (www.malaria.wr4.org)."

BASE COUNT 158 a 185 c 203 g 165 t
ORIGIN

Alignment Scores:
Pred. No.: 3,25e-20 Length: 711
Score: 263.00 Matches: 62
Percent Similarity: 47.96% Conservative: 32
Best Local Similarity: 31.63% Mismatches: 68
Query Match: 7.72% Indels: 34
DB: 12 Gaps: 6

US-09-843-007A-2 (1-636) x BM652118 (1-711)

Qy 118 GYAPLeuLYGlyLeuLYsAspLYsileProTYrPhe---GInGluLeuGlyLeuThr 136
Db 174 GGAGATTTCGGCGGATTAATGGAAAAGTTCCGTACTTCGACGCGAGCTCGGCATCGAT 233
Qy 137 TYrLeuHisLeuMetProLeuPheLYsCysProGluGlyLYsSerAspGlyTYrAla 156
Db 234 GCAATATGCTGTCGCGGATCTTCAAGTCACGG-----ATGGCGGACTTTGGCTACGAT 287
Qy 157 ValSerSerTYrArgAspValasnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
Db 288 ATTGCGGATTTCCGGGACATTCATTTCGGAGTTTGGAAACAATTGCTGACCTGGAGGCGTG 347
Qy 177 IleAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr 196
Db 348 GCGACCGCTTGTAAATGCAGAGGGGCTCAAGTANCTCGACTTTGTTCGACACCAACG 407
Qy 197 SerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216
Db 408 AGCGATGAAGTGGAGTGGTTTCTTAAAGTCCGTCAGAGGATCCCACTACGCGATTAC 467
Qy 217 TYrTYrIlePheProAspArg-----ArgMetProAspGlnTYrAsp 230
Db 468 TATGTGTCGATCCGGGTAAAGACGCTGGCTAACGGGACGCGGCTCT----- 515
Qy 231 ArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe---SerGlnLeuGlu 249

Db 516 -----CCTTCGAACTGGGTGAGTCTCTCCGT 542

Qy 250 AspGlyArgTIpValTIpThr -----ThrPheAsnSer 260
 Db 543 GGTTACGCTGGGAGTGAACGACGCTGCGGAAGGAGTACTATCTCCACCAAGTTCTAGTT 602

Qy 261 PheGlnTIpAspLeuAsnLeuGlyValAspIleLeuArgMetAspAlaVal 296
 Db 603 AACACGCCGATCTGAACATCTCCATCTCCGCTAGTACAGAAATGAAGAAGCTGATG 662

Qy 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaVal 296
 Db 663 ACGTTCTGCTGGCAAGGCGTCCATGATTCGGATCGATCGCGTG 710

RESULT 12
 BM628529 704 bp mRNA linear EST 26-FEB-2002
 LOCUS 17000687498686 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
 DEFINITION 19600449632435 5', mRNA sequence.

ACCESSION BM628529
 VERSION BM628529.1 GI:1828040
 SOURCE ESI.
 ORGANISM Anopheles gambiae (African malaria mosquito)
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
 Anopheles.

REFERENCE 1 (bases 1 to 704)
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gade Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU010049VT row: J column: 17
 Seq primer: M13 Reverse.

FEATURES
 Location/Qualifiers
 1..704
 /organism="Anopheles gambiae"
 /mol_type="mRNA"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone="19600449632435"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /clone_lib="A.Gam.ad.cDNA1"
 /notes="Vector: pSport1; Site 1: Sali; Site 2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen.
 cDNA inserts >500 bp cloned directionally into pSport 1.
 Not 1 site is 3'. Clones available through the Malasia
 Research and Reference Reagent Resource Center
 (www.malaria.mr4.org)."

BASE COUNT 160 a 162 c 213 g 169 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.67e-20 Length: 704
 Score: 262.50 Matches: 68
 Percent Similarity: 42.38% Conservative: 46
 Best Local Similarity: 25.28% Mismatches: 77
 Query Match: 7.70% Indels: 78
 DB: 12 Gaps: 10

US-09-843-007A-2 (1-636) x BM628529 (1-704)

Qy 59 TyrGlyAsnGlnLalaLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrrp--- 77
 Db 19 CATGGGAACATCCAAA-----CCATTCTGAGTGTGTGTCAGCCCTGTGCTGGGC 69

```

QY 78 -----GlnSerTyrSerGlnArgAsnSerSerLeuLysAspIle 90
Db 70 ACTCTTTAGTGGCATGTTAGATCTTAC----- 99
QY 91 AspIleAlaArgGluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyVal 110
Db 100 -----CGTACCAACGTCCTCG-----CGATTGGTGGGAAAGGC 132
QY 111 Cys-TyrValAspLeuPhe-----AlaGlyAs 119
Db 133 TGCCTTCTATCAATCTATCCCGATGCTTCATGACAGTGTGGGAGTGTGGTGGCGCA 192
QY 119 pLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHi 139
Db 193 TTTTAACGAATGCTTCGAAGTGGCTGACCTGAAGTGCATCGAGTGAAGCGCTTCTG 252
QY 139 sLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSe 159
Db 253 GATGTCACCGATCTACAAATCACC-----ATGTCGATTTGGCTATGATATTCTGA 306
QY 159 rTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAla 179
Db 307 CTTTCGGGACATCCAGAGAGTTTGGACGATGCTGATTTGACCGTTCGTGGAGCA 366
QY 179 aLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGl 199
Db 367 GCGCACGGGCTCGCTCGAAGGTGATCATGAGCTTTGTGCGGACCATTCGACAAATCT 426
QY 199 uHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTr 219
Db 427 GCACGAGTGGTTCGTAAGTCCGAACACGCGAGCCCGGCTATGAGAC---TATTACGT 483
QY 219 ePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAs 239
Db 484 GTGGCATGATGGAGGTT----- 502
QY 239 pGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAs 259
Db 503 ----AATCGCGCGGGGGTAGGAACITACCACAAATTCGATTCGAAGCCTTCGTGG 558
QY 259 nSer---PheGlnTrp----- 263
Db 559 CAGTGCATCGCATGGAGTGATCAGCGGCAACAGTACTACCTACACCATTCACCGTCGA 618
QY 264 -----AspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 281
Db 619 GCAACGGGACTTAACCTACGAAATCCGACCGTGTTCAGGAGATGAAGATGTCTGCT 678
QY 281 uPheLeuAlaAsnLeuGlyValAsp 289
Db 679 GTTCTGGCTGGCAAGGGCGTGGAC 703

RESULT 13
BM650445 706 bp mRNA linear EST 26-FEB-2002
LOCUS
DEFINITION
19600449650921 5', mRNA sequence.
ACCESSION
BM650445
VERSION
BM650445.1 GI:18949956
KEYWORDS
EST.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 706)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
JOURNAL
Unpublished
COMMENT
Contact: Holt R.A.
Celera Genomics

```

```

45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: Holt@celera.com
Plate: NU010049W7 row: L column: 23
Seq primer: M13 Reverse.
Location/Qualifiers
source
1..706
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/dev stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdNAL"
/notes="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."
BASE COUNT 160 a 162 c 215 g 169 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 3.69e-20 Length: 706
Score: 262.50 Matches: 68
Percent Similarity: 42.38% Conservative: 46
Best Local Similarity: 25.28% Mismatches: 77
Query Match: 7.70% Indels: 78
DB: 12 Gaps: 10

```

US-09-843-007A-2 (1-636) x BM650445 (1-706)

```

QY 59 TyrGlyAsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrp--- 77
Db 19 CATGGGACTCCAAA-----CCCATCTGATGTGTGTGTAGCCCTTGTGGGCG 69
QY 78 -----GlnSerTyrSerGlnArgAsnSerSerLeuLysAspIle 90
Db 70 ACTCTCTTTAGTGGCATGTTTAGATCTTAC----- 99
QY 91 AspIleAlaArgGluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyVal 110
Db 100 -----CGTACCAACGTCCTCG-----CGATTGGTGGGAAAGGC 132
QY 111 Cys-TyrValAspLeuPhe-----AlaGlyAs 119
Db 133 TGCCTTCTATCAATCTATCCCGATGCTTCATGACAGTGTGGGAGTGTGGGCGCA 192
QY 119 pLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHi 139
Db 193 TTTTAACGAATGCTTCGAAGTGGCTGACCTGAAGTGCATCGAGTGAAGCGCTTCTG 252
QY 139 sLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSe 159
Db 253 GATGTCACCGATCTACAAATCACC-----ATGTCGATTTGGCTATGATATTCTGA 306
QY 159 rTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAla 179
Db 307 CTTTCGGGACATCCAGAGAGTTTGGACGATGCTGATTTGAGCGGCTTGGTGGAGCA 366
QY 179 aLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGl 199
Db 367 GCGCACGGGCTCGCTCGAAGGTGATCATGAGCTTTGTGCGGACCATTCGACAAATCT 426
QY 199 uHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTr 219
Db 427 GCACGAGTGGTTCGTAAGTCCGAACACGCGAGCCCGGCTATGAGAC---TATTACGT 483
QY 219 ePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAs 239

```

RESULT 15	BM579790	694 bp	linear	EST 22-FEB-2002
LOCUS	BM579790			
DEFINITION	17000687241299 A.Gam.ad.cdna.blood1 Anopheles gambiae cdna clone			
ACCESSION	19600449716264 5', mRNA sequence.			
VERSION	BM579790			
KEYWORDS	BM579790.1 GI:18868257			
SOURCE	EST.			
ORGANISM	Anopheles gambiae (African malaria mosquito)			
	Anopheles gambiae			
	Eukaryota; Metazoa			
	Eukaryota; Arthropoda; Hexapoda; Insecta; Pterygota;			
	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;			
	Anopheles			
REFERENCE	1 (bases 1 to 694)			
AUTHORS	Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab			
	, R., Collins, P.H., Venter, J.C. and Hoffman, S.L.			

